ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE

ROBUSTE REGRESSION

THESE

PRESENTEE AU DEPARTEMENT DE MATHEMATIQUES DE
L'ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ES SCIENCES TECHNIQUES
PAR

Peter LISCHER
Mathématicien EPFZ
originaire de Willisau
(Lucerne)

acceptée sur proposition des
Prof. S. D. Chatterji, rapporteur
Prof. P. Nüesch
Dr. F. Streit, corapporteurs

Thèse No 160
Lausanne EPFL 1973

1. Einleitung

Wir gehen von einem Problem der klassischen Methode der kleinsten Quadrate aus : p unbekannte Parameter $\theta_1, \theta_2, \ldots, \theta_p$ sind aus n Beobachtungen y_i, $i = 1,2,\ldots, n$ zu schätzen. Wir nehmen an, dass

$$(1) \qquad y_i = \sum_{j=1}^{p} c_{ij}\,\theta_j + x_i \;, \; i = 1,2,\ldots,n$$

ist mit bekannten Koeffizienten c_{ij} und Fehlern x_i. Die Verteilungen der Fehler brauchen weder bekannt noch identisch zu sein. Man setzt aber voraus, dass sie Elemente einer gewissen Menge $\mathfrak{F}$ von Verteilungsfunktionen sind. Traditionellerweise lösen die Naturwissenschafter dieses Problem durch Minimalisierung der Summe der Quadrate

$$(2) \qquad \sum_{i=1}^{n} (y_i - c_i'\theta)^2$$

$c_i'\theta := \sum_{j=1}^{p} c_{ij}\theta_j$ bedeutet das Produkt der (1xp)-Matrix

$c_i' := (c_{i1},..,c_{ip})$ mit der (px1)-Matrix θ. ' bedeutet "transponiert", d.h. falls C eine (nxm)-Matrix ist, bedeutet C' diejenige Matrix, die aus C durch Vertauschung von Zeilen und Spalten hervorgeht.

Eicker [4], [5] hat notwendige und hinreichende Bedingungen angegeben, denen $\mathfrak{F}$ und die $\{c_{ij}\}$ genügen müssen, sodass die Methode der kleinsten Quadrate eine konsistente und asymptotisch normale Schätzung der Parameter $\theta_1, \theta_2, \ldots, \theta_p$ liefert. Die Methode der kleinsten Quadrate ist aber nur dann allen anderen Schätzmethoden vorzuziehen, wenn die Fehler normalverteilt und ihre Varianzen bekannt sind. Nun macht man aber häufig die Wahrnehmung, dass grosse

Fehler etwas zahlreicher auftreten, als nach dem Gauss'schen Fehlergesetz erwartet werden dürfte. Diese Tatsache wurde schon bei der ersten umfassenden Prüfung dieses Gesetzes von Bessel [1] hervorgehoben Es ist deshalb einleuchtend, dass bei einer quadratischen Verlustfunktion die entsprechenden schlechten Messungen ein zu grosses Gewicht erhalten. Um diesem Nachteil abzuhelfen , verwenden wir Schätzfunktionen $T_n = T_n\ (y_1, y_2, \ldots y_n)$ die den Ausdruck

$$(3) \qquad \sum_{i=1}^{n} \rho(y_i - c_i' T_n)$$

zu einem Minimum machen. $\rho(t)$ bedeutet hier eine nichtkonstante, meistens aber konvexe, reellwertige Funktion mit einem reellen Argument und $\rho(0) = 0$, z.B.

$$(4) \qquad \rho_k(t) := \begin{cases} \frac{1}{2}\, t^2 & \text{für } |t| \leqslant k \\ k\,|t| - \frac{1}{2}\, k^2 & \text{für } |t| > k \end{cases}$$

Der Wert k kann noch von den Beobachtungen y_i abhängen. Diese speziellen Verlustfunktionen liefern Schätzfunktionen mit gewissen wohldefinierten Minimaxeigenschaften, wenigstens im einfachsten Spezialfall ($p = 1$, $c_{ij} = 1\ \forall i,j$) [7].
Im 2. Abschnitt werden hinreichende Bedingungen angegeben, denen $\rho(t)$, $\mathfrak{F}$ und die $\{c_{ij}\}$ genügen müssen, damit die Schätzfunktion konsistent ist, die (3) zu einem Minimum macht. Wir verzichten auf grösstmögliche Allgemeinheit, da für robuste Schätzungen nur solche Verlustfunktionen $\rho(t)$ interessant sind, die für $t \to \pm\infty$ nicht zu stark ansteigen. Wir betrachten die folgenden Fälle :

(i) $\rho(t)$ konvex, gerade mit $\lim\limits_{t\to\infty} \dfrac{\rho(t)}{|t|} < \infty$

Voraussetzungen, die Konsistenz und asymptotische Normalität garantieren, haben den Vorteil, dass die Bedingungen über $\mathfrak{F}$, $\rho(t)$ und die Regressionskonstanten $\{c_{ij}\}$ getrennt angegeben werden können. Im Falle der asymptotischen Normalität wird dies durch einen Satz von Eicher [4] ermöglicht. Würde man den zentralen Grenzwertsatz in seiner üblichen Form verwenden, so erhielte man Bedingungen, die gleichzeitig die Regressionskonstanten und die Verteilungen der Fehler enthielten.

Für robuste Schätzungen sind die Verlustfunktionen (4) zweckmässig. Falls an den Stellen $\pm k$ die Verteilungen unstetig sind oder grosse Dichten aufweisen, ist es vorteilhaft, $\rho_k(t)$ so abzuändern, dass auch die 2. Ableitung stetig bleibt.

Wir haben in der ganzen Arbeit angenommen, dass die Regressionskonstanten $\{c_{ij}\}$ vorgegeben sind. Man könnte aber die Fragestellung abändern, nämlich : wie muss man bei gegebener Stichprobengrösse n die $\{c_{ij}\}$ wählen, dass einzelne Messungen nicht zu stark ins Gewicht fallen ? Ich vermute, dass sich mit Methoden, wie sie beim Beweis von Lemma 1 und 2 verwendet wurden, gewisse Faustregeln ableiten lassen sollten.

Beispiele für mögliche Verlustfunktionen $\rho(t)$ findet man in [7], [8] .

Diese Verlustfunktionen $\rho(t)$ lassen sich auf die meisten Beispiele anwenden, wie sie in Büchern über die Methode der kleinsten Quadrate zu finden sind. (z.B. Linnik [10])

2. Konsistenz

$(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitsraum, $\mathfrak{F}$ eine Menge von Verteilungsfunktionen, $x_1, x_2, \ldots$ unabhängige Zufallsgrössen auf $(\Omega, \mathfrak{A})$ mit Werten in $R = (-\infty, +\infty)$ und Verteilungen $F_1, F_2, \ldots \in \mathfrak{F}$ und $c_1, c_2, \ldots$ p-dimensionale Vektoren.

Wir betrachten Zufallsgrössen der Form

(1) $\quad y_i := c_i' \theta_o + x_i, \ \theta_o \in R^p \ , \ i = 1,2\ldots$

Gesucht ist eine Schätzung des Parameters θ_o.

In diesem Abschnitt untersuchen wir Schätzfunktionen $T_n := T_n(y_1, \ldots, y_n)$ für den Parameter θ_o, die den Ausdruck

(2) $$\sum_{i=1}^{n} \rho(y_i - c_i' T_n)$$

zu einem Minimum machen oder genauer :
Gesucht sind hinreichende Bedingungen dafür, dass jede Folge $T_n : R^n \to R^p$ mit der Eigenschaft

(3) $$\frac{1}{n} \sum_{i=1}^{n} \rho(y_i - c_i' T_n) - \inf_{\theta} \frac{1}{n} \sum_{i=1}^{n} \rho(y_i - c_i' \theta) \to 0 \quad \text{f.s.}$$

fast sicher gegen θ_o konvergiert. $\rho(t)$ bedeutet hier eine nicht-konstante Funktion, $\rho(t) \geq 0$, $\rho(0) = 0$.
Ohne Beschränkung der Allgemeinheit darf angenommen werden, dass $\mathfrak{F}$ keine Verteilung enthält, die die ganze Masse in einem Punkte konzentriert. Gäbe es nämlich ein $F \in \mathfrak{F}$ mit verschwindender Varianz, so bestünde eine lineare Beziehung zwischen den Parametern und die Anzahl der Parameter könnte um 1 reduziert werden.

2.1. Konvexe Verlustfunktionen

Sei $\rho(t)$ eine konvexe, reellwertige Funktion, definiert auf der reellen Achse, welche für $t \to \pm \infty$ nach $+ \infty$ strebt. $\psi(t) := \rho'(t)$ sei die normalisierte Ableitung von ρ, d.h. $\psi(t) = \frac{1}{2}[\psi(t+0) + \psi(t-0)]$. Wählen wir z.B. für $\rho(t) = |t|$, so erhalten wir den Ansatz von Laplace, welcher im Gegensatz zu Gauss die Methode der kleinsten Absolutbeträge postulierte.

Um die Voraussetzungen durchsichtiger zu machen, nehmen wir in diesem Abschnitt an, dass alle Fehler x_i die gleiche Verteilung F besitzen, d.h. dass $\mathfrak{F} = \{F\}$ ist. Für eine gegebene Verlustfunktion lassen sich leicht umfassendere Klassen angegeben, sodass unser Konsistenzbeweis gültig bleibt.

T_n erfülle (3). Wir beweisen Konvergenz von T_n gegen θ_o unter den folgenden Bedingungen :

(K1) $\rho(t)$ ist konvex und gerade ; $\rho(0) = 0$

(K2) $\lim\limits_{|t| \to \infty} \frac{\rho(t)}{|t|} = M_1 < \infty$

Sei $\gamma(t) := E\{\rho(x-t) - \rho(x)\}$

(K3) $\gamma(t) > \gamma(0) \quad \forall\ t \neq 0$

Sei $C_n := \frac{1}{\sqrt{n}}(c_{ir})$, $1 \leqslant i \leqslant n$, $1 \leqslant r \leqslant p$ und $D_n := C_n' C_n$.

(K4) (i) $|c_{ir}| \leqslant M_2 < \infty \ \forall i, r$

(ii) Es existiert eine Konstante $h > 0$ und eine Zahl $n_o < \infty$, so dass für alle $n \geqslant n_o$ $\det D_n \geqslant h$ ist.

Bemerkungen zu den Voraussetzungen :

(i) Aus (K1) und (K2) folgt mit Hilfe des motonen Konvergenzsatzes

$$(5) \qquad E\{\inf_{t' \in U} \rho(x-t') - \rho(x)\} \to E\{\rho(x-t) - \rho(x)\}$$

falls die Umgebung U von t auf $\{t\}$ zusammenschrumpft.

(ii) Gegeben sei eine Folge von Vektoren $c_1, c_2, \ldots,$ welche (K4) erfüllt und eine Zahl $n \geq n_o$. Wir wählen aus den n ersten Vektoren p linear unabhängige aus. Diese bilden eine Basis des p-dimensionalen Vektorraumes. Sofern möglich, wählen wir aus den restlichen (n-p) Vektoren wieder p linear unabhängige aus usw....

Um Einblick in die geometrische Bedeutung von (K4) zu erhalten, fragen wir uns nach der Anzahl Basen, die sich mit wachsendem n auf diese Weise konstruieren lassen.

Lemma 1

Falls (K4) erfüllt ist, existieren 2 Konstanten $K_1>0$ und $K_2>0$, sodass sich für alle $n \geq n_o$ aus den Vektoren $\{c_1, c_2, \ldots, c_n\}$ $\alpha(n) \geq K_1 n$ Basen $E_j := (e_j^{(1)}, e_j^{(2)}, \ldots, e_j^{(p)})$, $1 \leq j \leq \alpha(n)$, konstruieren lassen, die der Bedingung $|\det(e_j^{(1)}, e_j^{(2)}, \ldots, e_j^{(p)})| \geq K_2$ genügen.

Beweis : Für alle $n \geq p$ gilt die Beziehung

$$(6) \qquad \det D_n = \frac{1}{n^p} \sum_k \delta_k^2$$

wobei δ_k alle p-reihigen Minoren, die in der (nxp)-Matrix $\sqrt{n}\, C_n$ enthalten sind, also $\binom{n}{p}$ Minoren, durchläuft. (siehe Linnik [10], pag. 22 ff.)

Wegen $|c_{ir}| \leq M_2 \quad \forall\, i,r$, ist

$$(7) \qquad \delta_k^2 \leq (pM_2^2)^p \quad , 1 \leq k \leq \binom{n}{p}$$

Wir wählen K_2 so klein, (z.B. $K_2^2 = \frac{p!h}{2}$), sodass für alle n

$$(8) \quad \frac{1}{n^p} \sum_{k \in I} \delta_k^2 \leq \frac{h}{2} \quad , \qquad I := \{ k : 1 \leq k \leq \binom{n}{p} \ , \ |\delta_k| < K_2 \}$$

Wegen $\frac{1}{n^p} \sum_{k \varepsilon I} \delta_k^2 \geq \frac{h}{2}$ und $\delta_k^2 \leq (pM_2^2)^p$ existieren für alle $n \geq n_o$ mindestens $\frac{h}{2(pM_2^2)^p} n^p$ Minoren δ_k, deren Betrag grösser oder gleich K_2 ist. Die Kolonnenvektoren jeder solcher Determinante δ_k bilden eine vollständige Basis des p-dimensionalen Vektorraumes. Entfernt man p Vektoren aus der Menge $\{c_1, c_2, \ldots, c_n\}$, so verringert sich die Anzahl dieser Determinanten um höchstens

$$\binom{n}{p} - \binom{n-p}{p} < \frac{1}{p!} \left[n^p - (n-2p)^p\right] =$$

$$= \frac{1}{p!} \left[n - (n-2p)\right]\left[n^{p-1} + n^{p-2}(n-2p) + \ldots + (n-2p)^{p-1}\right] < \frac{2p^2}{p!} n^{p-1}$$

d.h. es lassen sich $\alpha(n) \geq \frac{hp!}{4p^2(pM_2^2)^p} \cdot n := K_1 n$ Basen auswählen, sodass jeder der Vektoren c_i in höchstens einer Basis verwendet wird und $|\det (e_i^{(1)}, e_j^{(2)}, \ldots, e_j^{(p)})| \geq K_2$ ist.

QED

(iii) Wie in [8] wird der Konsistenzbeweis in 2 Schritten durchgeführt : zuerst wird gezeigt, dass eine kompakte Menge $C \subset R^p$ existiert, sodass jede Folge, welche (3) erfüllt, fast sicher schliesslich in C bleibt; in einem zweiten Schritt, dass $T_n \to \theta_o$ f.s. Um die Beweisidee von [8] übernehmen zu können, braucht es folgende Ueberlegung : Gegeben sei ein Vektor $\theta \neq 0$. Wir bilden alle möglichen Skalarprodukte $c_1'\theta, c_2'\theta, \ldots, c_n'\theta$ und teilen sie in 2 Klassen ein, nämlich

$$I_n(\theta,K) := \{i : 1 \leq i \leq n, |c_i'\theta| \geq K\}$$

$$K>0$$

$$J_n(\theta,K) := \{i : 1 \leq i \leq n, |c_i'\theta| < K\}$$

$\|\theta\|$ bedeute die Euklidische Norm von θ und $|\theta|$ die sup-Norm, d.h. $|\theta| = \max (|\theta_1|, |\theta_2|, \ldots, |\theta_p|)$

Lemma 2

Falls (K4) erfüllt ist, existiert eine Konstante $b > 0$, sodass für alle $\theta \neq 0$ und alle $n \geq n_o$ $I_n(\theta, b\|\theta\|)$ mindestens $\alpha(n) \geq K_1 n$ Elemente enthält.

Bew : Wir betrachten alle Matrizen $E_j := (e_j^{(1)}, e_j^{(2)}, \ldots, e_j^{(p)})$ mit $|\det E_j| \geq K_2$, $1 \leq j \leq \alpha(n)$ und bilden alle Skalarprodukte $c_r'\theta$ von θ mit den Spaltenvektoren c_r der Matrizen E_j , $1 \leq r \leq p$.

Ferner sei v_j (resp. w_j) der kleinste (resp. grösste) Eigenwert der positiv definiten Matrix $E_j'E_j$. Es gilt

$$(9) \qquad w_j \|\theta\|^2 \geq \|E_j'\theta\|^2 = \sum_{r=1}^{p} |c_r'\theta|^2 \geq v_j \|\theta\|^2$$

und

$$(10) \qquad v_j \geq \frac{K_2}{w_j^{p-1}} \geq p\, b^2 \qquad 1 \leq j \leq \alpha(n)$$

für ein $b > 0$, d.h. in jeder Basis E_j mit $|\det E_j| \geq K_2$ existiert mindestens ein Basisvektor c_r so, dass

$$(11) \qquad |c_r'\theta| \geq b\|\theta\|$$

QED

Lemma 3

Unter den Voraussetzungen (K1) bis (K4) existiert eine kompakte Menge $C \subset R^p$ so, dass jede Folge T_n, welche (3) erfüllt, fast sicher schliesslich in C bleibt.

Bew : Zu zeigen ist, dass eine kompakte Menge $C \subset R^p$ existiert mit $\theta_o \in C$, sodass für alle genügend grossen n fast sicher

$$(12) \qquad \inf_{\theta \notin C} \frac{1}{n} \sum_{i=1}^{n} [\rho(y_i - c_i'\theta) - \rho(y_i - c_i'\theta_o)] \geq \text{const.} > 0$$

ist. Es gilt wegen (18)

$$(13) \qquad E\,\{\liminf_{|t| \to \infty} \frac{\rho(x-t) - \rho(x)}{\rho(t) + 1}\} \geq 1$$

Es ist aber für alle i $\inf_{\theta \in C} [\rho(y_i - c_i'\theta) - \rho(y_i - c_i'\theta_o)] \leq 0$.

Um (13) gleichwohl verwenden zu können, muss man die Lemmata 1 und 2 zu Hilfe nehmen.

Aus den Vektoren $\{c_1, c_2, \ldots, c_n\}$ lassen sich nämlich Basen $E_j := (e_j^{(1)}, \ldots, e_j^{(p)})$, $1 \leq j \leq \alpha(n)$, konstruieren, die der Bedingung $|\det E_j| \geq K_2$ genügen.

Sei $I_j(n) := \{i : 1 \leq i \leq n,\ c_i \varepsilon E_j\}$

$$b(t) := \rho(t) + 1 \ ; \quad b_j(\theta,n) := \sum_{s \varepsilon I_j(n)} b(c_s'(\theta-\theta_o))$$

$$A_j(\theta,n) := \sum_{s \varepsilon I_j(n)} \frac{\rho(y_s - c_s'\theta) - \rho(y_s - c_s'\theta_o)}{b_j(\theta,n)}; \ 1 \leq j \leq \alpha(n)$$

Jeden der übrigen Vektoren $c_k \neq 0$ ergänzen wir zu einer orthonomierten Basis

$F_k := (\frac{c_k}{\|c_k\|} := f_k^{(1)}, f_k^{(2)}, \ldots, f_k^{(p)})$ mit passenden $f_k^{(2)}, \ldots, f_k^{(p)}$.

Sei $b_k'(\theta,n) := \sum_{s=1}^{p} b(M_2 f_k^{(s)'}(\theta-\theta_o))$ und

$$A'_k\ (\theta,n) := \frac{\rho(y_k - c'_k\theta) - \rho(y_k - c'_k\theta_o)}{b'_k(\theta,n)}\ ,\ k \notin \bigcup_{j=1}^{\alpha(n)} I_j(n)\ ,\quad k \leqslant n$$

Sei $C_t \subset R^p$ mit $\theta_o \varepsilon C_t$ eine kompakte Menge derart, dass jede Basis $E := (e^{(1)}, e^{(2)}, .., e^{(p)})$ mit $|\det E| \geqslant K_2$ sowie

$$\max_{1 \leqslant s \leqslant p} |e^{(s)}| \leqslant M_2$$

$$\inf_{\theta \notin C_t} \sum_{s=1}^{p} |e^{(s)'}(\theta - \theta_o)| > t \text{ ist.}$$

Wir zeigen : es existiert eine Zahl t_o derart, dass für ein beliebiges ε mit $0 < \varepsilon < pK_1$

(14) $$E\{\inf_{\theta \notin C_{t_o}} A_j(\theta,n)\} > 1 - \frac{\varepsilon}{2}\quad ,\ 1 \leqslant j \leqslant \alpha(n)$$

(15) $$E\{\inf_{\theta \notin C_{t_o}} A'_k(\theta,n)\} > -\frac{\varepsilon}{2p}\quad ,\ k \notin \bigcup_{j=1}^{\alpha(n)} I_j(n)\quad ,\quad k \leqslant n$$

(16) $$\inf_{\theta \notin C_{t_o}} b_j(\theta,n) \geqslant \frac{1}{K_1}\quad ,\ 1 \leqslant j \leqslant \alpha(n)$$

Wegen des Mittelwertsatzes und wegen $\rho(x) \leqslant M_1\ |x|$ gilt

(17) $$|\rho(x-t) - \rho(x)| \leqslant M_1\ |t|$$

(18) $$\rho(x-t) - \rho(t) \geqslant \rho(x) - 2M_1\ |x|$$

Ferner gilt

(19) $$\liminf_{|\theta| \to \infty} b_j(\theta,n) = \infty,\quad 1 \leqslant j \leqslant \alpha(n)$$

Wegen (18) lässt sich ein t_1 so wählen, dass für alle $|x| \leq a$, für alle j, $1 \leq j \leq \alpha(n)$ und beliebiges ε, $0 < \varepsilon < pK_1$, (wir setzen x_s für $y_s - c_s'\theta_o$)

$$(20) \quad \inf_{\theta \notin C_{t_1}} A_j(\theta,n) \geq \inf_{\theta \notin C_{t_1}} \sum_{s \in I_j(n)} \frac{\rho(c_s'(\theta-\theta_o))-2M_1|x_s|}{b_j(\theta,n)} \geq 1-\frac{\varepsilon}{6}$$

Da für alle genügend grossen v $\rho(v) \geq M_1|\frac{v}{2}|$ ist, gilt wegen (17) für alle x und alle genügend grossen t_2

$$(21) \quad \inf_{\theta \notin C_{t_2}} A_j(\theta,n) \geq \inf_{\theta \notin C_{t_2}} \sum_{s \in I_j(n)} - \frac{M_1 \; |c_s'(\theta-\theta_o)|}{b_j(\theta,n)} \geq -2p$$

Somit gilt für ein a mit der Eigenschaft $|F(a) - F(-a)|^p \geq 1-\frac{\varepsilon}{12p}$ und $t_3 := \max(t_1,t_2)$

$$(22) \quad E\{\inf_{\theta \notin C_{t_3}} A_j(\theta,n)\} \geq (1-\tfrac{\varepsilon}{6})(1-\tfrac{\varepsilon}{12p})-2p\;\{1-[F(a)-F(-a)]^p\} \geq$$

$$\geq (1-\tfrac{\varepsilon}{6})(1-\tfrac{\varepsilon}{12p})-2p\;\frac{\varepsilon}{12p} > 1-\frac{\varepsilon}{2}$$

Ferner existiert eine Zahl t_4 so, dass für alle $k \notin \bigcup_{j=1}^{\alpha(n)} I_j(n)$, $k \leq n$

$$(23) \quad E\{\inf_{\theta \notin C_{t_4}} A_k'(\theta,n)\} \geq \int_{-a}^{+a} \inf_{\theta \notin C_{t_4}} \frac{\rho(c_k'(\theta-\theta_o)) - 2M_1 \; |x_k|}{b_k'(\theta,n)} F(dx) +$$

$$+ \inf_{\theta \notin C_{t_4}} \{- \frac{M \; |c_k'(\theta-\theta_o)|}{b_k'(\theta,n)}\}(1-F(a)+F(-a)) \geq$$

$$\geqslant \inf_{\theta \notin C_{t_4}} \left\{ - \frac{2M_1 a}{b_k'(\theta,n)} \right\} (F(a)-F(-a)) +$$

$$+ \inf_{\theta \notin C_{t_4}} \left\{ - \frac{M_1 |c_k'(\theta-\theta_o)|}{b_k'(\theta,n)} \right\} (1-F(a) + F(-a)) \geqslant$$

$$\geqslant - \frac{\varepsilon}{4p} \left(1 - \frac{\varepsilon}{12p}\right) - \frac{2\varepsilon}{12p} > - \frac{\varepsilon}{2p}$$

Schliesslich existiert eine Zahl t_5 so, dass

$$(24) \qquad \inf_{\theta \notin C_{t_5}} b_j(\theta,n) \geqslant \frac{1}{K_1}, \quad 1 \leqslant j \leqslant \alpha(n)$$

und somit auch

$$(25) \qquad \inf_{\theta \notin C_{t_5}} b_k'(\theta,n) \geqslant \frac{1}{K_1}, \quad k \notin \bigcup_{j=1}^{\alpha(n)} I_j(n), \quad k \leqslant n$$

Wegen des starken Gesetzes der grossen Zahlen gilt für $t_o := \max(t_3,t_4,t_5)$ und alle genügend grossen n fast sicher.

$$(26) \quad \inf_{\theta \notin C_{t_o}} \left\{ \frac{1}{n} \sum_{j=1}^{\alpha(n)} A_j(\theta,n) + \frac{1}{n} \sum_k A_k'(\theta,n) \right\} \geqslant$$

$$\geqslant \frac{1}{n} \sum_{j=1}^{\alpha(n)} \inf_{\theta \notin C_{t_o}} A_j(\theta,n) + \frac{1}{n} \sum_k \inf_{\theta \notin C_{t_o}} A_k'(\theta,n) \geqslant$$

$$\geqslant K_1(1-\varepsilon) - (1 - pK_1) \frac{\varepsilon}{p} = K_1 - \frac{\varepsilon}{p}$$

Also ist

$$(27) \quad \frac{1}{n}\sum_{i=1}^{n} [\rho(y_i - c_i'\theta) - \rho(y_i - c_i'\theta_o)] \geq (K_1 - \frac{\varepsilon}{p}) \frac{1}{K_1} = 1 - \frac{\varepsilon}{K_1 p} > 0$$

für alle $\theta \notin C_{t_o}$, womit das Lemma bewiesen ist, denn für alle n gilt

$$(28) \qquad \inf_\theta \frac{1}{n}\sum_{i=1}^{n} [\rho(y_i - c_i'\theta) - \rho(y_i - c_i'\theta_o)] \leq 0$$

QED

Bemerkung :

Im 3. Abschnitt benützen wir dieses Lemma in einer allgemeineren Form : Seien s', s'' 2 beliebige Konstanten, $0 < s' \leq s'' < \infty$ und $(s_n)_{n \geq 1} \subset [s', s'']$ eine beliebige Folge von Zahlen.

Beh : Unter den Voraussetzungen (K1) bis (K4) existiert eine kompakte Menge $C_o \subset R^p$ und eine Nullmenge N, sodass jede Folge $T_n(s_n)$, welche

$$(3') \quad \sum_{i=1}^{n} \rho\left|\frac{y_i - c_i' T_n(s_n)}{s_n}\right| - \inf_\theta \sum_{i=1}^{n} \rho\left|\frac{y_i - c_i'\theta}{s_n}\right| = 0$$

erfüllt, für alle $\omega \notin N$ schliesslich in C_o bleibt.

Der Beweis bleibt sich gleich. Ersetzen wir nämlich die x_i und θ durch $\frac{x_i}{s}$ und $\frac{\theta}{s}$, so folgt unmittelbar die Existenz einer kompakten Menge $C_o \subset R^p$ mit der Eigenschaft :

Für ein beliebiges ε, $0 < \varepsilon < pK_1$ und alle genügend grossen n gilt f.s.

$$(29)\qquad \inf_{\substack{\theta \notin C_o \\ s\varepsilon[s',s'']}} \frac{1}{n}\sum_{i=1}^{n}\left[\rho\left[\frac{y_i - c_i'\theta}{s}\right] - \rho\left[\frac{y_i - c_i'\theta_o}{s}\right]\right] \geq 1 - \frac{\varepsilon}{pK_1} > 0$$

womit die Behauptung bewiesen ist, denn für alle n gilt

$$(30)\qquad \inf_{\substack{\theta \\ s\varepsilon[s',s'']}} \frac{1}{n}\sum_{i=1}^{n}\left[\rho\left[\frac{y_i - c_i'\theta}{s}\right] - \rho\left[\frac{y_i - c_i'\theta_o}{s}\right]\right] \leq 0$$

Satz 1

Unter den Voraussetzungen (K1) bis (K4) konvergiert jede Folge T_n, welche (3) erfüllt, fast sicher gegen θ_o.

Bew : Wir können uns auf die kompakte Menge C beschränken. Sei U eine offene Umgebung von θ_o. Wegen (K3) und (K4) (siehe Lemma 2) ist

$\frac{1}{n}\sum_{i=1}^{n} \gamma(c_i'(\theta-\theta_o))$ für alle $\theta\varepsilon C\setminus U$ und alle $n \geq n_o$ strikt grösser als Null, das heisst

$$(31)\qquad \frac{1}{n}\sum_{i=1}^{n} \gamma(c_i'(\theta-\theta_o)) \geq 3\varepsilon$$

für ein $\varepsilon > 0$. $\rho(x-t)$ ist stetig in t. Zu jedem $\theta\varepsilon C\setminus U$ gehört eine Umgebung U_θ, sodass für alle i wegen (5)

$$(32)\qquad E\{\inf_{\hat\theta\varepsilon U_\theta} \rho(y_i - c_i'\hat\theta) - \rho(y_i - c_i'\theta_o)\} \geq \gamma(c_i'(\theta-\theta_o)) - \varepsilon$$

Also ist

$$(33)\quad \frac{1}{n}\sum_{i=1}^{n} E\{\inf_{\hat\theta\varepsilon U_\theta} \rho(y_i-c_i'\hat\theta)-\rho(y_i-c_i'\theta_o)\} \geq \frac{1}{n}\sum_{i=1}^{n} \gamma(c_i'(\theta-\theta_o))-\varepsilon \geq 2\varepsilon$$

Da $C \setminus U$ kompakt ist, lässt sich eine endliche Ueberdeckung U_s, $1 \leq s \leq N$, auswählen. Wegen des starken Gesetzes der grossen Zahlen gilt fast sicher für alle genügend grossen n und alle $1 \leq s \leq N$

$$(34)\quad \inf_{\hat\theta\varepsilon U_s} \frac{1}{n}\sum_{i=1}^{n} [\rho(y_i-c_i'\hat\theta)-\rho(y_i-c_i'\theta_o)] \geq$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \inf_{\hat\theta\varepsilon U_s} [\rho(y_i-c_i'\hat\theta)-\rho(y_i-c_i'\theta_o)] \geq$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} E\{\inf_{\hat\theta\varepsilon U_s} \rho(y_i-c_i'\hat\theta)-\rho(y_i-c_i'\theta_o)\} - \varepsilon \geq$$

$$\geq \frac{1}{n}\sum_{i=1}^{n} \gamma(c_i'(\theta-\theta_o))-2\varepsilon \geq \varepsilon$$

womit der Satz bewiesen ist, denn es gilt für alle n

$$(35)\quad \inf_{\theta\varepsilon U} \frac{1}{n}\sum_{i=1}^{n} [\rho(y_i-c_i'\theta)-\rho(y_i-c_i'\theta_o)] \leq 0$$

QED

2.2. Gestutzte Verlustfunktionen

Sei $\rho(t)$ eine konvexe, nicht-konstante, gerade und reellwertige Funktion, definiert auf der reellen Achse.

Wir betrachten folgende Verlustfunktion :

$$\rho_m(t) := \begin{cases} \rho(t) & , |t| \leqslant m \\ \rho(m) & , |t| > m \end{cases} , m > 0$$

Wir nehmen wiederum $\mathfrak{F} := \{F\}$ an.

Gesucht sind hinreichende Bedingungen dafür, dass jede Folge

$T_n := T_n(y_1, y_2, \ldots, y_n)$, welche

$$(*) \quad \frac{1}{n} \sum_{i=1}^{n} \rho_m(y_i - c_i' T_n) - \inf_{\theta} \frac{1}{n} \sum_{i=1}^{n} \rho_m(y_i - c_i' \theta) \to 0 \text{ fast sicher}$$

für ein festes $m > 0$ erfüllt, fast sicher gegen θ_o konvergiert.

Satz 2

Vor :

(i) Es existiert eine Zahl $m_o > 0$ so, dass für alle festen $m \varepsilon [m_o, \infty[$ und alle $t \neq 0$ $E\{\rho_m(x-t)\} > E\{\rho_m(x)\}$ ist.

(ii) Die Vektoren $\{c_i\}$ erfüllen (K4)

Beh :

Es existiert eine Zahl m_1, $m_o \leqslant m_1 < \infty$, so dass jede Folge $T_n : R^n \to R^p$, welche (*) erfüllt, für alle festen $m \varepsilon [m_1, \infty[$ fast sicher gegen θ_o konvergiert.

Bew :

Wir zeigen zuerst, dass zu jedem genügend grossen m eine kompakte Menge $C_m \subset R^p$ existiert, sodass T_n schliesslich fast sicher in C_m bleibt. Aus Satz 1 folgt dann unmittelbar $T_n \to \theta_o$ fast sicher.
(Im Beweis von Satz 1 wurde die Konvexität nicht verwendet)
Es gilt für alle festen m > 0

$$(36) \qquad E\{\liminf_{|t|\to\infty} \frac{\rho_m(x-t)}{\rho(m)}\} = 1$$

und

$$(37) \qquad \lim_{m\to\infty} \frac{E\{\rho_m(x)\}}{\rho(m)} = 0$$

Beweis von (37): Sei $\delta > 0$ gegeben. Es gilt für ein beliebiges a > 0

$$(38) \quad \frac{E\{\rho_m(x)\}}{\rho(m)} = \int_0^a \frac{\rho_m(x)}{\rho(m)} F(dx) + \int_a^\infty \frac{\rho_m(x)}{\rho(m)} F(dx) + \int_{-a}^0 \frac{\rho_m(x)}{\rho(m)} F(dx) + \int_{-\infty}^{-a} \frac{\rho_m(x)}{\rho(m)} F(dx)$$

Wir wählen ein $a < \infty$ so, dass $[1-F(a)+F(-a)] < \frac{\delta}{2}$ und ein $\hat{m}$ so, dass $\frac{\rho_{\hat{m}}(a)}{\rho(\hat{m})} < \frac{\delta}{2}$ ist. Dann gilt

$$(39) \quad \frac{\rho_{\hat{m}}(a)}{\rho(\hat{m})} < \frac{\delta}{2} (F(a)-F(-a))+1-F(a)+F(-a) < \delta$$

Da δ beliebig klein gewählt werden kann, ist (37) bewiesen.
$I_j(n)$ habe die gleiche Bedeutung wie in 2.1. Sei

$$A_j(\theta,m,n) := \sum_{s\in I_j(n)} \frac{\rho_m(y_s - c_s'\theta)}{\rho(m)}, \quad j = 1,2,\dots, \alpha(n)$$

Zu jedem festen $m > 0$ existiert eine kompakte Menge $C_m \subset R^p$ so, dass

(40) $$E\{\inf_{\theta \notin C_m} A_j(\theta,m,n)\} \geq 1 - \frac{\varepsilon}{4}, \quad j = 1,2,\ldots,\alpha(n)$$

ist für ein $\varepsilon]0,1[$. Wir wählen $m_1 \geq m_o$ so gross, dass

(41) $$\frac{E\{\rho_{m_1}(x)\} + \varepsilon}{\rho(m_1)} \leq K_1(1-\varepsilon)$$

Wegen des starken Gesetzes der grossen Zahlen gilt fast sicher für alle genügend grossen n

(42) $$\inf_{\theta \notin C_{m_1}} \frac{1}{n} \sum_{i=1}^{n} \frac{\rho_{m_1}(y_i - c_i'\theta)}{\rho(m_1)} \geq \frac{1}{n} \sum_{j=1}^{\alpha(n)} \inf_{\theta \notin C_{m_1}} A_j(\theta,m_1,n) \geq K(1-\frac{\varepsilon}{2})$$

Für alle genügend grossen n gilt aber fast sicher

(43) $$\inf_{\theta} \frac{1}{n} \sum_{i=1}^{n} \frac{\rho_{m_1}(y_i - c_i'\theta)}{\rho(m_1)} \leq \frac{1}{n} \sum_{i=1}^{n} \frac{\rho_{m_1}(y_i - c_i'\theta_o)}{\rho(m_1)} \leq$$

$$\leq \frac{E\{\rho_{m_1}(x)\} + \varepsilon}{\rho(m_1)} \leq K_1(1-\varepsilon)$$

QED

3. Gleichzeitige Schätzung eines Regressions- und eines Massstabparameters

Wir betrachten das Regressionsmodell aus Abschnitt 2.1. und suchen wiederum eine Schätzung T_n für den Regressionsparameter. P.J. Huber [7] untersuchte das Problem, robuste Schätzungen eines Translationsparameters zu bestimmen, falls die zugehörige wahre Verteilung aus der Menge

$\{F:F(t):=(1-\varepsilon)\phi(\frac{t}{\sigma})+\varepsilon H(t)\}$ stammt. $\phi(t):=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\exp(-\frac{x^2}{2})dx$,

ε,σ unbekannt, $H(t)$ symmetrisch.

Von den 3 Lösungsideen scheint der 2. Ansatz sich am zweckmässigsten anwenden zu lassen, auch wenn die wahren Verteilungen nicht der oben erwähnten Menge angehören. Ferner besitzen Schätzungen, die nach dieser Methode erhalten werden, aequivalente asymptotische Eigenschaften mit dem gestutzten Mittelwert ("trimmed mean") [2].
Aus praktischen Gründen wird man aber den gestutzten Mittelwert verwenden, um einen Translationsparameter zu schätzen, während die andere Methode bei Regressionsproblemen vorzuziehen ist.
Die Verallgemeinerung des Ansatzes von Huber auf Regressionsparameter läuft auf die Lösung des Gleichungssystems

$$(1)\qquad \begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} c_{ir}\,\psi\left[\frac{y_i - c_i'T_n}{S_n}\right] = 0 && r = 1,2,\ldots,p \\ & && i = 1,2,\ldots,n \\ &\frac{1}{n}\sum_{i=1}^{n} \psi^2\left[\frac{y_i - c_i'T_n}{S_n}\right] = \beta,\ \beta > 0;\ n = 1,2,\ldots\ldots \end{aligned}$$

hinaus. ψ: $R\to R$ bedeutet dabei eine monoton wachsende Funktion, z.B. $\psi(t) := \max(-k, \min(k,t))$, $k > 0$.

Ueber die Wahl von β siehe [7]. Wir nehmen hier lediglich an, dass $E\{\psi^2(\frac{x}{s})\} = \beta$ eine eindeutige Lösung für $s = s_o$ besitzt. Wir werden fast sichere Konvergenz von (T_n, S_n) gegen $(\theta_o, s_o)'$ beweisen. Für $c_{ij} = 1 \; \forall i,j; p = 1$ folgt zusätzlich aus diesem Resultat, (Satz 3), dass der 2. Ansatz von Huber([7], pag 96) wirklich eine konsistente Schätzung des Translations- und des Massstabparameters liefert.

Ohne Beweis sei angeführt, dass sowohl $\sqrt{n}\,(T_n - \theta_o)$, als auch $\sqrt{n}\,(S_n - s_o)$ unter sehr allgemeinen Bedingungen asymptotisch normalverteilt sind. Dies lässt sich mit Hilfe der Methoden aus [8] und Abschnitt 4 nachweisen.
Fragen der Existenz und Eindeutigkeit der Lösung von (1) werden am Schluss des Abschnittes untersucht.

3.1. Konsistenz

$(\Omega,\mathfrak{A},P)$ sei ein Wahrscheinlichkeitsraum, $\mathfrak{F}$ eine Menge von stetigen symmetrischen Verteilungsfunktionen [d.h. $\forall F \varepsilon \mathfrak{F}$ gilt $F(x) = 1 - F(-x) \; \forall x \varepsilon R$], $x_1, x_2, \ldots$ unabhängige Zufallsgrössen auf $(\Omega,\mathfrak{A})$ mit Werten in $R := (-\infty,+\infty)$ und Verteilungen $F_1, F_2, \ldots \varepsilon \mathfrak{F}$ und $c_1, c_2, \ldots$ p-dimensionale Vektoren.
$\rho : R \to R$ sei eine gerade, stetig differenzierbare und konvexe Funktion mit der Ableitung $\rho' = \psi$ und $\rho(0) = 0$
Wir betrachten Zufallsgrössen der Form $y_i = c_i'\theta_o + x_i$, $\theta_o \varepsilon R^p$, $i = 1,2,\ldots$ (T_n, S_n) sei für jedes n eine Lösung des Gleichungssystems (1).
Wir nehmen wieder der Einfachheit wegen $\mathfrak{F} = \{F\}$ an.

Satz 3

Vor :

(i) $\psi(t)$ ist stetig, ungerade, monoton und in einer Umgebung des Nullpunktes strikt monoton wachsend.

(ii) $\lim\limits_{t \to +\infty} \psi(t) = M_1 < \infty$

(iii) $\lambda(t) := E\{\psi(x-t)\} = 0$ genau dann, wenn $t = 0$ ist

(iv) $E\{\psi^2(\frac{x}{s})\} = \beta$ besitzt eine eindeutig bestimmte Lösung für $s = s_o$.

(v) Die Vektoren $\{c_i\}$ genügen der Voraussetzung (K4), d.h.

(i) $|c_{ir}| \leqslant M_2 < \infty \ \forall i, r$

(ii) Es existiert eine Konstante $h > 0$ und eine Zahl $n_o < \infty$ so, dass für alle $n \geqslant n_o$ $\det D_n \geqslant h$ ist.

Beh :

Jede Folge (T_n, S_n), welche (1) erfüllt, konvergiert fast sicher gegen (θ_o, s_o), wobei s_o durch die Gleichung $E\{\psi^2(\frac{x}{s_o})\} = \beta$ bestimmt wird.

Bew :

Wir zeigen zuerst, dass 2 Zahlen s' und s'', $0 < s' \leqslant s'' < \infty$, existieren, sodass (T_n, S_n) fast sicher schliesslich in $R^p x [s', s'']$ bleibt.
Das Koordinatensystem sei so gewählt, dass $\theta_o = 0$ wird.

(i) Wir wählen $s' > 0$ so klein, dass für alle positiven $s \leq s'$

$$\inf_{t\in R} E\{\psi^2(\frac{x+t}{s})\} \geq \beta + b$$

ist für ein $b > 0$.
Wegen der Voraussetzung (iv), S. 22 gilt nämlich $\forall s \in]0,s_o[$

$M_1^2 \geq E\{\psi^2(\frac{x}{s})\} > \beta$. Wählt man $\delta > 0$ so, dass

$$\sup_{t\in R} \{F(t+\delta)-F(t-\delta)\} \leq \frac{M_1^2 - \beta}{4M_1^2}$$

und $s' > 0$ so, dass $\psi^2(\frac{\delta}{s'}) \geq \beta + \frac{M_1^2 - \beta}{2}$, so erhält man unmittelbar die Umgleichung

$$\forall t\in R, \forall s\in]0,s']: E\{\psi^2(\frac{x+t}{s})\} \geq [\beta + \frac{M_1^2-\beta}{2}]\,[1 - \frac{M_1^2-\beta}{4M_1^2}] > \beta + \frac{M_1^2-\beta}{4} := \beta + b$$

Für alle festen $s \in]0,s']$ und alle festen $\theta \in R$ gilt

$$(2)\quad \beta+b \leq \inf_{t\in R} E\{\psi^2(\frac{x+t}{s})\} \leq \frac{1}{n}\sum_{i=1}^{n} E\{\psi^2\left[\frac{x_i - c_i'\theta}{s}\right]\},\ n = 1,2,\ldots$$

Sei $N(\theta,s) := \{\omega: \limsup_{n\to\infty} \left|\frac{1}{n}\sum_{i=1}^{n} \psi^2\left[\frac{x_i - c_i'\theta}{s}\right] - E\{\psi^2\left[\frac{x_i - c_i'\theta}{s}\right]\}\right| > 0\}$

Wegen des starken Gesetzes der grossen Zahlen ist $N(\theta,s)$ für alle festen s und alle festen θ eine Nullmenge. $\theta^{(r)}$, $r=1,2\ldots$ seien diejenigen $\theta \in R^p$, deren Komponenten alle rational sind.

$$f(t,\omega) := \limsup_{n\to\infty} \left|\frac{1}{n}\sum_{i=1}^{n} [\psi^2(x_i(\omega)-t) - E\{\psi^2(x_i(\omega)-t)\}]\right|, t\in R, \omega\in\Omega$$

ist für festes ω eine gleichmässig stetige Funktion in t.

Es gilt für alle $\omega \notin \bigcup_{r=1}^{\infty} N(\theta^{(r)},s)$:

$$(3)\quad \limsup_{n\to\infty} \left|\frac{1}{n}\sum_{i=1}^{n} \psi^2\left[\frac{x_i - c_i'\theta^{(r)}}{s}\right] - E\{\psi^2\left[\frac{x_i - c_i'\theta^{(r)}}{s}\right]\}\right| = 0 \;\forall\, \theta^{(r)}, r=1,2,\ldots$$

Da die linke Seite von (3) stetig in θ ist, gilt $\forall \omega \notin \bigcup_{r=1}^{\infty} N(\theta^{(r)},s)$:

$$(4)\quad \limsup_{n\to\infty} \left|\frac{1}{n}\sum_{i=1}^{n} \psi^2\left[\frac{x_i - c_i'\theta}{s}\right] - E\{\psi^2\left[\frac{x_i - c_i'\theta}{s}\right]\}\right| = 0, \quad \forall\, \theta\varepsilon R^p$$

d.h. aber : für alle genügend grossen n und für alle $s\varepsilon]0,s']$ gilt fast sicher

$$(5)\qquad \frac{1}{n}\sum_{i=1}^{n} [\psi^2\left[\frac{x_i - c_i'\theta}{s}\right] - \beta] \geqslant \frac{b}{2}, \qquad \forall\, \theta\varepsilon R^p$$

(ii) Es existiert eine Zahl $q<\infty$ so, dass $P\{\,|x_i|>q\} \leqslant \frac{\varepsilon}{2}$ ist für ein $\varepsilon > 0$. Wegen der gleichmässigen Stetigkeit von $\psi(t)$ lässt sich ein s'' so wählen, dass für alle $s \geqslant s''$ und alle $\theta\varepsilon R^p$ sowohl

$$(6)\quad \left|\frac{1}{n}\sum{}' c_i\psi\left[\frac{x_i - c_i'\theta}{s}\right] - \frac{1}{n}\sum{}' c_i\psi\left[-\frac{c_i'\theta}{s}\right]\right| < \varepsilon$$

als auch

$$(7)\quad \left|\frac{1}{n}\sum{}' \psi^2\left[\frac{x_i - c_i'\theta}{s}\right] - \frac{1}{n}\sum{}' \psi^2\left[-\frac{c_i'\theta}{s}\right]\right| < \varepsilon$$

$\sum'$ (bzw. $\sum''$) bedeutet Summation über diejenigen Indizes i, für die $|x_i| \leqslant q$ (bzw. $|x_i|>q$) ist. $(1 \leqslant i \leqslant n)$

Sei $C_1 := \{(\theta,s): s>s'', \; \|\frac{\theta}{s}\| \leqslant \delta\}$, $\;C_2 := \{(\theta,s): s>s'', \; \|\frac{\theta}{s}\| >\delta\}$

$\delta > 0$ sei so gewählt, dass für alle n, alle θ mit $\|\theta\| \leq \delta$ und alle c_i mit $|c_i| \leq M_2$

$$\frac{1}{n} \sum_{i=1}^{n} \psi^2(-c_i'\theta) < \frac{\beta}{2}$$

ist. Für alle genügend grossen n gilt fast sicher

$$(8) \quad \frac{1}{n} \sum{}' \psi^2 \left[\frac{x_i - c_i'\theta}{s}\right] + \frac{1}{n} \sum{}'' \psi^2 \left[\frac{x_i - c_i'\theta}{s}\right] \leq$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \psi^2 \left[- \frac{c_i'\theta}{s}\right] + \varepsilon + M_1^2 \varepsilon \quad , \forall (\theta, s) \varepsilon C_1$$

$$(9) \quad \left|\frac{1}{n} \sum{}' c_i \psi \left[\frac{x_i - c_i'\theta}{s}\right] + \frac{1}{n} \sum{}'' c_i \psi \left[\frac{x_i - c_i'\theta}{s}\right]\right| \geq$$

$$\geq \left|\frac{1}{n} \sum_{i=1}^{n} c_i \psi \left[- \frac{c_i'\theta}{s}\right]\right| - \varepsilon - 2M_1M_2\varepsilon \quad , \forall (\theta, s) \varepsilon C_2$$

Wegen (K4) existiert eine Konstante $a > 0$, sodass für alle $n \geq n_o$

$$(10) \quad \left|\frac{1}{n} \sum_{i=1}^{n} c_i \, \psi \left[- \frac{c_i'\theta}{s}\right]\right| \geq a \quad , \qquad \forall (\theta, s) \varepsilon C_2$$

<u>Bew :</u> $G_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} \rho(c_i'\theta)$ ist konvex. Da $\lim_{|t| \to \infty} \frac{\rho(t)}{|t|} = M_1$

ist, existiert wegen Lemma 2 eine Konstante d, sodass für alle $n \geq n_o$

$$\inf_{\|\theta\| > d} \frac{G_n(\theta)}{\|\theta\|} > \frac{bK_1M_1}{2} \text{ und überdies } \inf_{\delta \leq \|\theta\| \leq d} G_n(\theta) \geq \text{const.} > 0$$

ist. Eine auf einer konvexen Menge $B \subset R^p$ definierte, stetig differenzierbare konvexe Funktion $G_n(\theta)$ erfüllt für alle $\tau \varepsilon R^p$, $\sigma \varepsilon R^p$ die Ungleichung

$$(11) \qquad G_n(\tau) \geq G_n(\sigma) + (\tau-\sigma)' \operatorname{grad} G_n(\sigma)$$

d.h. für $\tau = 0$, $\sigma = \frac{\theta}{s}$, $G_n(0) = 0$ gilt

$$(12) \qquad - \frac{\theta'}{s} \frac{1}{n} \sum_{i=1}^{n} c_i \psi\left(-\frac{c_i'\theta}{s}\right) \geq G_n\left(\frac{\theta}{s}\right)$$

Für $(\theta,s) \varepsilon C_2$ existiert somit eine Konstante $a > 0$, sodass

$G_n\left(\frac{\theta}{s}\right) \geq a \cdot \left\|\frac{\theta}{s}\right\|$ ist, womit (10) bewiesen ist.

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, existiert ein $\varepsilon_1 > 0$, sodass für alle genügend grossen n fast sicher die rechte Seite von (8) kleiner als $\frac{3}{4}\beta$ und diejenige von (9) grösser als $\frac{a}{2}$ bleibt.

(iii) Nach der Bemerkung nach Lemma (3) können wir uns nun auf die kompakte Menge $C := C_o x [s', s''] \subset R^{p+1}$ beschränken. Sei

$$\Psi\left[\frac{x_i - c_i'\theta}{s}\right] := \begin{pmatrix} c_i \; \psi\left[\frac{x_i - c_i'\theta}{s}\right] \\ \\ \psi^2 \left[\frac{x_i - c_i'\theta}{s}\right] - \beta \end{pmatrix}$$

und

$$\Lambda(c_i'\theta, s) := E\{\Psi\left[\frac{x_i - c_i'\theta}{s}\right]\}$$

U sei eine offene Umgebung von $(0, s_o)$.

Wegen (iii), (iv) und (v) der Voraussetzung ist

$$\left|\frac{1}{n}\sum_{i=1}^{n} \Lambda(c_i'\theta,s)\right|$$

für alle genügend grossen n strikt positiv aufder kompakten Menge $C \setminus U$, das heisst

(13) $$\left|\frac{1}{n}\sum_{i=1}^{n} \Lambda(c_i'\theta,s)\right| \geq 5\varepsilon > 0$$

Dies beweist man analog (10). Zuerst zeigt man, dass für jede offene Umgebung U_o des Nullpunktes, alle $s\varepsilon[s',s'']$ und alle $n \geq n_o$

$\left|\frac{1}{n}\sum_{i=1}^{n} c_i E\{\psi\left[\frac{x_i-c_i'\theta}{s}\right]\}\right| \geq$ const. > 0 ist. Nach Voraussetzung (iv) besitzt aber $E\{\psi^2\left[\frac{x_i}{s}\right]\} - \beta$ eine eindeutige Nullstelle für $s = s_o$.

Somit leistet jede beliebige offene Umgebung U von $(0,s_o)$ das Verlangte.

Für jedes $(\theta,s) \in C\setminus U$ sei $U_{\theta,s}$ eine Umgebung von (θ,s) derart, dass für alle i

(14) $$E\left\{\sup_{(\hat\theta,\hat s)\varepsilon U_{\theta,s}} \left|\psi\left[\frac{x_i-c_i'\hat\theta}{\hat s}\right] - \psi\left[\frac{x_i-c_i'\theta}{s}\right]\right|\right\} \leq \varepsilon$$

und folglich auch

(15) $$\left|\Lambda(c_i'\hat\theta,s) - \Lambda(c_i'\theta,s)\right| \leq \varepsilon$$

ist für alle $(\hat\theta,\hat s)\in U_{\theta,s}$.

Wir wählen eine endliche Ueberdeckung U_r : U_{θ_r, s_r}, $1 \leq r \leq N$

Für alle genügend grossen n gilt f.s.

$$(16) \qquad \sup_{(\hat\theta,\hat s)\in C\setminus U} \left| \frac{1}{n} \sum_{i=1}^{n} \left[\psi\left[\frac{x_i - c_i'\hat\theta}{\hat s}\right] - \Lambda(c_i'\hat\theta,\hat s)\right]\right| \leq$$

$$\leq \sup_{1\leq r\leq N} \frac{1}{n} \sum_{i=1}^{n} \sup_{(\hat\theta,\hat s)\in U_r} \left|\psi\left[\frac{x_i - c_i'\hat\theta}{\hat s}\right] - \psi\left[\frac{x_i - c_i'\theta_r}{s_r}\right]\right| +$$

$$+ \sup_{1\leq r\leq N} \left| \frac{1}{n} \sum_{i=1}^{n} \left[\psi\left[\frac{x_i - c_i'\theta_r}{s_r}\right] - \Lambda(c_i'\theta_r, s_r)\right]\right| +$$

$$+ \left| \frac{1}{n} \sum_{i=1}^{n} \Lambda(c_i'\theta_r, s_r) - \frac{1}{n} \sum_{i=1}^{n} \Lambda(c_i'\hat\theta,\hat s)\right| \leq 4\varepsilon$$

Da $\left|\frac{1}{n} \sum_{i=1}^{n} \Lambda(c_i'\theta, s)\right| \geq 5\varepsilon$ ist für $(\theta,s)\in C\setminus U$, impliziert das

$$(17) \qquad \left|\frac{1}{n} \sum_{i=1}^{n} \psi\left[\frac{x_i - c_i'\theta}{s}\right]\right| \geq \varepsilon$$

für alle $\theta\in C\setminus U$ und alle genügend grossen n, womit der Satz bewiesen ist.

3.2. Endliche Stichproben

Wir zeigen jetzt, dass (1) eine legitime Definition von T_n und S_n ergibt. Sei $y_1, y_2, \ldots, y_n$ eine feste Stichprobe von Umfang n. Es ist Existenz und Eindeutigkeit der Lösung des Gleichungssystems (1) zu zeigen. Wir führen den Beweis für die spezielle Funktion $\psi(t) := \psi(t,k) := \max(-k,\min(k,t))$, $k>0$.

Mit analogen, allerdings unübersichtlicheren Voraussetzungen lässt sich der Beweis auch für allgemeinere Funktionen führen.

<u>Satz 4</u>

Sei (β,k) gegeben mit $\beta < k^2$, D_n nicht-singulär und die Stichprobe $y_1,y_2,\ldots,y_n$ habe die Eigenschaft (*) :

Aus $y_{i_1} - c_{i_1}'\theta = y_{i_2} - c_{i_2}'\theta = \ldots = y_{i_m} - c_{i_m}'\theta$, $i_1 < i_2 < \ldots < i_m$,

$\theta \in R^p$, $\{i_1,i_2,\ldots,i_m\} \subset \{1,2,\ldots,n\}$

folgt $m < n(1 - \frac{\beta}{k^2})$

Dann besitzt das Gleichungssystem (1) eine Lösung (T_n,S_n) mit einem eindeutig bestimmten S_n.

Ist zusätzlich $(\hat{D}_n)_{rs} := \frac{1}{n} \sum\limits_I c_{ir}c_{is}$, $1 \leq r \leq p$, $1 \leq s \leq p$

$I := \{i: -kS_n < y_i - c_i'T_n < kS_n\}$, nicht-singulär, so ist auch

T_n eindeutig bestimmt.

<u>Bemerkung :</u> Falls $F_i = \mathcal{L}(x_i)$, $1 \leq i \leq n$, stetig ist, so ist die Eigenschaft (*) mit Wahrscheinlichkeit 1 erfüllt.

<u>Bew :</u>

(i) Für jedes Paar (T,S), welches $\frac{1}{n} \sum\limits_{i=1}^{n} \psi^2 \left[\frac{y_i - c_i'T}{S}\right] = \beta$ erfüllt, ist die Anzahl n_2 derjenigen y_i, für die $|y_i - c_i'T| < kS$ ist, grösser oder gleich $n(1 - \frac{\beta}{k^2})$, da

$$(n - n_2)k^2 \leq \sum_{i=1}^{n} \psi^2 \left[\frac{y_i - c_i'T}{S}\right] \tag{18}$$

Zu jedem festen $s > 0$ existiert mindestens eine Lösung $T(s)$ des Systems

$$(19) \qquad \sum_{i=1}^{n} c_i \; \psi\left[\frac{y_i - c_i' T(s)}{s}\right] = 0 \, .$$

Sei $\quad Z_s := \{\theta : \sum_{i=1}^{n} c_i \; \psi\left[\frac{y_i - c_i'\theta}{s}\right] = 0\} = \{\theta : \sum_{i=1}^{n} \rho\left[\frac{y_i - c_i'}{s}\right] = \min.\}$

Z_s ist konvex und kompakt. Sei $\hat{s} < \infty$.

<u>Beh :</u> $\qquad B := \bigcup_{0 < s \leq \hat{s}} Z_s$ ist beschränkt.

<u>Bew :</u> Annahme : B sei nicht beschränkt.

Es gilt für alle $T(s) \varepsilon B$

$$(20) \quad \frac{1}{n} \sum_{i=1}^{n} \rho\left[\frac{y_i - c_i' T(s)}{s}\right] \leq \frac{1}{n} \sum_{i=1}^{n} \rho\left[\frac{y_i}{s}\right] < \frac{1}{n} \sum_{i=1}^{n} k \left|\frac{y_i}{s}\right|$$

Es lässt sich ein $T(s_1) \in B$ so gross wählen, dass für mindestens ein $j \varepsilon \{1,2,\ldots,n\}$

$$\rho\left[\frac{y_j - c_j' T(s_1)}{s_1}\right] > \sum_{i=1}^{n} k \left|\frac{y_i}{s_1}\right| .$$

ist, was im Widerspruch zu (20) steht.

QED

Nach (*) sind Bindungen von hoher Multiplizität ausgeschlossen. Deshalb existiert ein $s' > 0$ und ein $T(s') \varepsilon Z_{s'}$, sodass

$$(21) \qquad \frac{1}{n} \sum_{i=1}^{n} \psi^2 \left[\frac{y_i - c_i' T(s')}{s'}\right] - \beta > 0$$

(Falls kein solches s' existierte, müsste es mindestens $n(1-\frac{\beta}{k^2})$ Messungen geben mit der Eigenschaft $|y_i - c_i'T(s)| < ks$ für alle $s > 0$ und wegen der Beschränktheit von B ein Vektor T mit der Eigenschaft $y_i - c_i'T = 0$ für ebenfalls $n(1-\frac{\beta}{k^2})$ Messungen, was im Widerspruch zu (*) steht.)

Es existiert ferner ein $s'' < \infty$ sodass für ein $T(s'') \in Z_{s''}$

$$(22) \qquad \frac{1}{n}\sum_{i=1}^{n} \psi^2\left[\frac{y_i - c_i'T(s'')}{s''}\right] - \beta < 0$$

Das stetige Bild einer zusammenhängenden Menge $W \subset R^{p+1}$ ist ebenfalls zusammenhängend. Die einzigen zusammenhägenden Mengen in R^1 sind Intervalle. Um die Existenz einer Lösung $(\tilde{s}, T(\tilde{s}))$ von (1) nachzuweisen, genügt es deshalb zu zeigen, dass $W := \bigcup_{s' \leq s \leq s''} s \times Z_s$ zusammenhängend ist.

Seien U und V disjunkte offene Mengen, sodass $W \subset U \cup V$, $U \cap V \cap W = \phi$ (ϕ bedeute die leere Menge) und

$\mathcal{S}_1 := \{s : s' \leq s \leq s'', \ s \times Z_s \subset U\}$, $\qquad \mathcal{S}_2 := \{s : s' \leq s \leq s'', \ s \times Z_s \subset V\}$

Zu zeigen: $\mathcal{S}_1$ und $\mathcal{S}_2$ sind abgeschlossen; denn aus $\mathcal{S}_1 \cup \mathcal{S}_2 = [s', s'']$, $\mathcal{S}_1 \cap \mathcal{S}_2 = \phi$, $\mathcal{S}_1 = \overline{\mathcal{S}}_1$, $\mathcal{S}_2 = \overline{\mathcal{S}}_2$ folgt entweder $\mathcal{S}_1 = [s', s'']$, $\mathcal{S}_2 = \phi$ oder $\mathcal{S}_1 = \phi$, $\mathcal{S}_2 = [s', s'']$.

Wir betrachten eine Folge $(s_i, t_i) \subset U$, $i = 1, 2, \ldots$ $s_i \to s^o$ für $i \to \infty$ Wegen der Beschränktheit von W existiert eine konvergente Teilfolge $(\hat{s}_j, \hat{t}_j) \to (s^o, t^o)$, $j \to \infty$.

Um $(s^o, t^o) \subset U$ nachzuweisen, benötigen wir noch die Stetigkeit von

$$M(s) := \min_{\theta} \frac{1}{n}\sum_{i=1}^{n} \rho\left[\frac{y_i - c_i'\theta}{s}\right] \quad \text{im Intervall } [s', s''].$$

Als Abkürzung benützen wir im folgenden $f(\theta, s) := \frac{1}{n}\sum_{i=1}^{n} \rho\left[\frac{y_i - c_i'\theta}{s}\right]$

$C_o \subset R^p$ sei eine kompakte Menge, sodass $\bigcup_{s' \leq s \leq s'} Z_s \subset C_o$, also

$W \subset W_1 := [s', s''] \times C_o$ ist.

$f(\theta, s)$ ist stetig auf W_1, d.h. $\forall\, \varepsilon > 0\; \exists\, \delta > 0$ sodass $\forall\, (\theta, s) \in W_1$ mit $|s - s^o| < \delta$ $|f(\theta, s) - f(\theta, s^o)| < \varepsilon$ ist.

$\forall\, (\theta, s) \in W$ mit $|s - s^o| < \delta$ gilt $f(\theta, s) \geqslant f(\theta, s^o) - \varepsilon > M(s^o) - \varepsilon$, also

(23) $$M(s) \geqslant M(s^o) - \varepsilon$$

$\forall\, (\theta^o, s^o) \in W$ mit $|s - s^o| < \delta$ gilt $M(s) \leqslant f(\theta^o, s) \leqslant f(\theta^o, s^o) + \varepsilon = M(s^o) + \varepsilon$ also

(24) $$M(s) \leqslant M(s^o) + \varepsilon$$

Aus (23) und (24) folgt unmittelbar die Stetigkeit von $M(s)$.

Wegen der Stetigkeit von $M(s)$ in $[s', s'']$ gilt $M(\hat{s}_j) \to M(s^o)$ für $\hat{s}_j \to s^o$.

Andererseits ist $\lim\limits_{j \to \infty} M(\hat{s}_j) = \lim\limits_{j \to \infty} f(\hat{s}_j, \hat{t}_j) = f(s^o, t^o)$, d.h. $(s^o, t^o) \in W$

Da das Komplement V^c abgeschlossen ist und $(\hat{s}_j, \hat{t}_j) \in V^c\ \forall\, j$, gilt auch $(s^o, t^o) \in V^c$, also $(s^o, t^o) \in U$, $s^o \in \mathcal{S}_1$.

Somit ist $\mathcal{S}_1 = \overline{\mathcal{S}}_1$. Analog zeigt man $\mathcal{S}_2 = \overline{\mathcal{S}}_2$.

(ii) Seien $(S_{(1)}, T_{(1)})$, $(S_{(2)}, T_{(2)})$ 2 Lösungen von (1), $S_{(1)} \neq S_{(2)}$.

Wir setzen $T(S) := T_{(1)} + d(S - S_{(1)})$ mit $d := \dfrac{T_{(2)} - T_{(1)}}{S_{(2)} - S_{(1)}}$

und betrachten die Funktion

$$(25) \quad R(S) := \sum_{i=1}^{n} \left[\frac{1}{2}\,\psi^2\left[\frac{y_i - c_i'T(S)}{S}\right] + d'c_i\,\psi\left[\frac{y_i - c_i'T(S)}{S}\right]\right] - \frac{n}{2}\,\beta$$

welche für $S = S_{(1)}$ und $S = S_{(2)}$ verschwindet. Die Ableitung $R'(S)$ lässt sich in der Form

$$(26) \quad R'(S) = -\frac{1}{S} \sum_{I} \left[d'\,c_i + \frac{y_i - c_i'T(S)}{S}\right]^2$$

schreiben mit $I := \{i : -kS < y_i - c_i'T(S) < kS\}$.

Da $R'(S)$ für $S_{(1)} \leq S \leq S_{(2)}$ das gleiche Vorzeichen besitzt, und $R(S_{(1)}) = R(S_{(2)}) = 0$ ist muss $R'(S)$ identisch verschwinden. Das heisst aber, dass $y_i = c_i'(T_{(1)} - S_{(1)}d)$ ist für alle $i \in I$, was eine Bindung der Multiplizität $m \geq n(1 - \frac{\beta}{k^2})$ bedeuten würde, entgegen der Voraussetzung (*).

Somit ist die Zahl S_n eindeutig bestimmt.

Der letzte Teil der Behauptung ergibt sich unmittelbar aus (19), denn $T_n(S_n)$ genügt der Beziehung

$$(27) \quad \hat{D}_n\,T_n(S_n) = \frac{1}{n}\left[\sum{}' c_i y_i + \sum{}'' c_i k S_n \operatorname{sgn}\left(y_i - c_i'T_n(S_n)\right)\right]$$

$\sum'$ (bzw. $\sum''$) bedeutet Summation über $i \in I$ (bzw. $i \notin I$)

QED

4. Asymptotische Normalität

$(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitsraum, $\mathfrak{F}$ eine Mengen von Verteilungsfunktionen, $x_1, x_2, \ldots$ unabhängige Zufallsgrössen auf $(\Omega, \mathfrak{A})$ mit Werten in R und Verteilungen $F_1, F_2, \ldots \varepsilon \mathfrak{F}$ und $c_1, c_2, \ldots$ p-dimensionale Vektoren.
$\psi: R \to R$ sei eine nicht-konstante Funktion.
Gesucht sind hinreichende Bedingungen dafür, dass jede Folge $T_n : R^n \to R^p$ mit der Eigenschaft

$$(1) \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_i \, \psi(y_i - c_i' T_n) \to 0 \text{ in Wahrscheinlichkeit}$$

asymptotisch normal ist.

Vorausgesetzt, die Konsistenz und die Messbarkeit von T_n seien schon durch irgendwelche Methoden gezeigt worden, wird asymptotische Normalität unter den folgenden Voraussetzungen besiesen :

(N1) (i) $\psi(t)$ ist Borel-messbar.

(ii) Es existiert eine abzählbare Teilmenge $S := \{s_1, s_2, \ldots\} \subset R$ so, dass für jedes feste $t \varepsilon R$ eine Teilfolge $\{s_{n_j}\} \subset S$ existiert mit der Eigenschaft

$$(2) \qquad \liminf_{s_{n_j} \to t} \psi(s_{n_j}) \leqslant \psi(t) \leqslant \limsup_{s_{n_j} \to t} \psi(s_{n_j})$$

Diese Vorausetzungssichert die Existenz der im folgenden vorkommenden Limites, Infima und Suprema. (vgl. Loève [11] pag. 505 ff).

Im ganzen Abschnitt bedeutet $|\theta|$ die sup-Norm, d.h.

$|\theta| := \max(|\theta_1|, |\theta_2|, \ldots, |\theta_p|)$

Sei $\lambda_i(t) := E\{\psi(x_i - t)\}$

$$Z_{ir}(\theta) := c_{ir}\,\psi(y_i - c_i'\,), \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant r \leqslant p$$

$$Z_i(\theta) := (Z_{i1}(\theta), Z_{i2}(\theta), \ldots, Z_{ip}(\theta))'$$

$$u(x_i, \sigma, c_i, d) := \sup_{|\tau - \sigma| \leqslant d} |\psi(x_i - c_i'\tau) - \psi(x_i - c_i'\sigma)|$$

(N2) $\lambda_i(0) := E\{\psi(x_i)\} = 0$ für alle x_i mit $\mathcal{L}(x_i) \varepsilon \mathfrak{F}$

(N3) Es existieren strikt positive Zahlen a_1, a_2, b, c, d_o so, dass für alle x_i mit $\mathcal{L}(x_i) \varepsilon \mathfrak{F}$

(i) $\lambda_i(t) := \int_{-\infty}^{+\infty} \psi(x_i - t)\, F_i(dx_i)$ ist in einer Umgebung des Punktes $t = 0$ stetig differenzierbar und es gilt ferner $a_1 \leqslant -\lambda_i'(0) \leqslant a_2$.

(ii) $E\{u(x_i, \sigma, c_i, d)\} \leqslant bd$ für $|\sigma| + d \leqslant d_o$, $d > 0$, $|c_i| \leqslant M$

(iii) $E\{u^2(x_i, \sigma, c_i, d)\} \leqslant cd$ für $|\sigma| + d \leqslant d_o$, $d > 0$, $|c_i| \leqslant M$

(N4)(i) $\sup_{F \varepsilon \mathfrak{F}} \int_{|x| > t} \psi^2(x) F(dx) \to 0$, falls $t \to +\infty$

(ii) $\inf_{F \varepsilon \mathfrak{F}} \int_{-\infty}^{+\infty} \psi^2(x) F(dx) > 0$

Sei $(D_n)_{rs} := \frac{1}{n} \sum_{i=1}^{n} c_{ir}\, c_{is}$, $\quad r, s = 1, 2, \ldots, p$

(N5)(i) $\quad |c_{ir}| \leq M < \infty \quad \forall i,r$

(ii) Es existiert eine Konstante $h > 0$ und eine Zahl $n_o < \infty$ so, dass für alle $n \geq n_o$ $\det D_n \geq h$ ist.

Bevor wir mit dem Beweis des Satzes anfangen, führen wir noch einige neue Bezeichnungen ein :

I_p bedeute die p-dimensionale Einheitsmatrix

$C_n := \frac{1}{\sqrt{n}} (c_{ir}),\ 1 \leq i \leq n,\ 1 \leq r \leq p$

$\Sigma_n := \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ mit $\sigma_k^2 := E\{\psi^2(x_k)\},\ 1 \leq k \leq n$

$\hat{\Lambda}_n = \operatorname{diag}(-\lambda_1'(0), -\lambda_2'(0), \ldots, -\lambda_n'(0))$ mit $\lambda_k'(0) = \frac{d}{dt} E\{\psi(x_k - t)\}|_{t=0}$

$D_n := C_n' C_n$

$\Lambda_n := C_n' \Sigma_n C_n$

$A_n^2 := C_n' \Sigma_n C_n = \operatorname{Cov}[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_i \psi(x_i)]$

<u>Bemerkung :</u>

Wegen (N3) (i) und (N5) gilt für alle $n \geq n_o$

(3) $$\det \Lambda_n := \det (C_n' \hat{\Lambda}_n C_n) \geq h\, a_1^p$$

Ueberdies gilt

(4) $$\frac{\partial}{\partial \theta_k} \frac{1}{n} \sum_{i=1}^{n} c_{ir} \lambda_i(c_i'(\theta - \theta_o))|_{\theta = \theta_o} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} c_{ir} c_{ik} \lambda_i'(0) = (-\Lambda_n)_{rk}$$

$\frac{1}{n}\sum_{i=1}^{n} c_i \lambda_i(c_i'(\theta-\theta))$ besitzt somit eine nicht-singuläre Ableitung $-\Lambda_n$ im Punkte $\theta=\theta_o$, d.h.

$$(5) \quad \left|\frac{1}{n}\sum_{i=1}^{n} c_i \lambda_i(c_i'(\theta-\theta_o)) + \Lambda_n(\theta-\theta_o)\right| = o(|\theta-\theta_o|)$$

Sei v_n der kleinste Eigenwert von Λ_n. Wegen (3) und weil die Matrixelemente von Λ_n beschränkt sind, existiert eine Konstante $a > 0$, sodass

$$(6) \qquad |v_n| \geq 2a > 0 \ , \ \forall\, n \geq n_o$$

Somit existiert eine Zahl $g > 0$, sodass für alle $n \geq n_o$

$$(7) \quad \left|\frac{1}{n}\sum_{i=1}^{n} c_i\lambda_i(c_i'(\theta-\theta_o))\right| = \left|\frac{1}{n}\sum_{i=1}^{n} E\{Z_i(\theta)\}\right| \geq a|\theta-\theta_o| \text{ für } |\theta-\theta_o| \leq g$$

$$\text{Sei} \quad W_{rn}(\tau,\sigma) := \frac{\sum_{i=1}^{n} |Z_{ir}(\tau)-E\{Z_{ir}(\tau)\}- Z_{ir}(\sigma) + E\{Z_{ir}(\sigma)\}|}{\sqrt{n} + \left|\frac{1}{n}\sum_{i=1}^{n} E\{Z_i(\tau)\}\right| \sum_{i=1}^{n} |c_{ir}|}$$

und $W_n(\tau,\sigma)' := (W_{1n}(\tau,\sigma), W_{2n}(\tau,\sigma), \ldots, W_{pn}(\tau,\sigma))$

Lemma 4

Vor : (N1) bis (N5)

Beh :

(8) $\sup\limits_{|\tau-\theta_o|\leq\bar{d}_o} |W_n(\tau,\theta_o)| \rightarrow 0$ in Wahrscheinlichkeit, falls $n\rightarrow\infty$

$\bar{d}_o := \min (d_o, g)$

Bew : Das Koordinatensystem sei so gewählt, dass $\theta_o = 0$ und $\bar{d}_o := \min (g, d_o) = 1$ wird. Die Idee des Beweises beruht darin, den Würfel $|\tau| \leq 1$ in eine wachsende Anzahl kleinerer Würfel zu unterteilen und $W_n(\tau,0)$ in jedem dieser kleinerer Würfel in Wahrscheinlichkeit beschränkt zu halten.

Wir setzen $q = \frac{1}{L}$, wobei $L \geq 2$ eine ganze Zahl ist, die später gewählt wird und betrachten die konzentrischen Würfel

(9) $K_s := \{\theta: |\theta| \leq (1-q)^s\}$, $s = 0,1,2,\ldots,s_o$

Wir unterteilen die Differenz $K_{s-1} \setminus K_s$ in kleinere Würfel (s. Figur 1) mit Kanten der Länge

(10) $$2d = (1-q)^{s-1}q$$

sodass die Koordinaten ihrer Mittelpunkte ξ ungerade Vielfache von d sind und

(11) $$|\xi| = (1-q)^s \; (1-\tfrac{q}{2})$$

ist. Zu jedem Wert von s gibt es weniger als $(2L)^p$ solcher kleinerer Würfel, sodass schliesslich $N < s_o(2L)^p$ Würfel in $K_o \setminus K_{s_o}$ enthalten sind; wir numerieren sie $K_{(1)}, K_{(2)}, \ldots, K_{(N)}$.

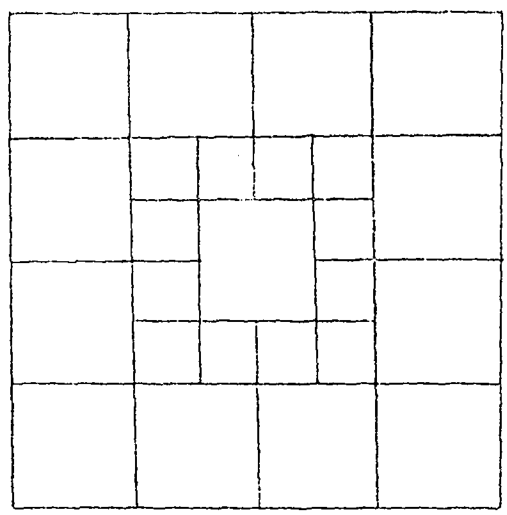

Fig. 1

Sei $\varepsilon > 0$ gegeben. Wir werden zeigen, dass für eine geeignete Wahl von L und $s_o = s_o(n)$ die rechte Seite von

$$(12) \quad P\{\sup_{\tau\varepsilon K_o} |W_n(\tau,0)| \geq 2\varepsilon\} \leq P\{\sup_{\tau\varepsilon K_{s_o}} |W_n(\tau,0)| \geq 2\varepsilon\} + \sum_{j=1}^{N} P\{\sup_{\tau\varepsilon K_{(j)}} |W_n(\tau,0)| \geq 2\varepsilon\}$$

mit wachsendem n gegen Null strebt, womit Lemma 4 bewiesen sein wird. Wir wählen

$$(13) \qquad L \geq \frac{3b}{\varepsilon a}$$

(siehe (7) und $s_o = s_o(n)$ ist definiert durch

$$(14) \qquad (1-q)^{s_o} \leq n^{-\gamma} < (1-q)^{s_o - 1}, \quad \gamma \text{ fest}, \ \tfrac{1}{2} < \gamma < 1$$

Somit ist

$$(15) \qquad N = O(\log n)$$

Nun nehmen wir irgendeinen der Würfel mit Zentrum ξ und Kantenlänge 2d in Uebereinstimmung mit (10), (11).
Für $\tau \varepsilon K_{(j)}$ haben wir wegen (7) und (N3)(ii)

$$(16) \qquad \left|\frac{1}{n}\sum_{i=1}^{n} E\{Z_i(\tau)\}\right| \geq a|\tau| \geq a(1-q)^s$$

$$(17) \quad |E\{\psi(x_i - c_i'\tau)\} - E\{\psi(x_i - c_i'\xi)\}| \leq E\{u(x_i,\xi,c_i,d)\} \leq bd \leq b(1-q)^s q$$

Es gilt

$$(18) \quad |W_{rn}(\tau,0)| \leq |W_{rn}(\tau,\xi)| + \frac{\left|\sum_{i=1}^{n} [Z_{ir}(\xi) - Z_{ir}(0) - E\{Z_{ir}(\xi)\}]\right|}{\sqrt{n} + \left|\frac{1}{n}\sum_{i=1}^{n} E\{Z_i(\tau)\}\right| \sum_{i=1}^{n} |c_{ir}|}$$

Somit ist

$$(19) \qquad \sup_{\tau \varepsilon K_{(j)}} |W_{rn}(\tau,0)| \leq U_{rn} + V_{rn}$$

mit

$$(20) \qquad U_{rn} := \frac{\sum_{i=1}^{n} |c_{ir}| [u(x_i,\xi,c_i,d) + E\{u(x_i,\xi,c_i,d)\}]}{a(1-q)^s \sum_{i=1}^{n} |c_{ir}|}$$

$$(21)\quad V_{rn} := \frac{\left|\sum_{i=1}^{n} |c_{ir}| \; [\psi(x_i - c_i'\xi) - \psi(x_i) - E\{\psi(x_i - c_i'\xi)\}]\right|}{a(1-q)^s \sum_{i=1}^{n} |c_{ir}|}$$

Wegen (13) und (17) gilt nun

$$(22)\quad P\{U_{rn} \geq \varepsilon\} =$$

$$= P\{\sum_{i=1}^{n} |c_{ir}| \; [u(x_i,\xi,c_i,d) + E\{u(x_i,\xi,c_i,d)\}] \geq \varepsilon a(1-q)^s \sum_{i=1}^{n} |c_{ir}|\}$$

$$= P\{\sum_{i=1}^{n} |c_{ir}|(u[i] - E\{u[i]\}) \geq \varepsilon a(1-q)^s \sum_{i=1}^{n} |c_{ir}| - 2\sum_{i=1}^{n} |c_{ir}| E\{u[i]\}\} \leq$$

$$\leq P\{\sum_{i=1}^{n} |c_{ir}|(u(x_i,\xi,c_i,d) - E\{u(x_i,\xi,c_i,d)\}) \geq b\, q(1-q)^s \sum_{i=1}^{n} |c_{ir}|\}$$

(N3) (iii) und die Ungleichung von Tschebyscheff ergeben

$$(23)\quad P\{U_{rn} \geq \varepsilon\} \leq \frac{c}{b^2 q(1-q)} \; \frac{\sum_{i=1}^{n} c_{ir}^2}{[\sum_{i=1}^{n} |c_{ir}|]^2} \; \frac{1}{(1-q)^{s-1}}$$

In ähnlicher Weise

$$(24)\quad P\{V_{rn} \geq \varepsilon\} \quad =$$

$$= P\{\sum_{i=1}^{n} |c_{ir}|\,|\psi(x_i - c_i'\xi) - \psi(x_i) - \lambda_i(c_i'\xi)| \geq \varepsilon a(1-q)^{s} \sum_{i=1}^{n} |c_{ir}|\} \leq$$

$$\leq P\{\sum_{i=1}^{n} |c_{ir}|\,u(x_o,\xi,c_i,d) \geq \varepsilon a(1-q)^{s} \sum_{i=1}^{n} |c_{ir}|\} \leq$$

$$\leq \frac{c}{9b^2q^2(1-q)^2} \; \frac{\sum_{i=1}^{n} c_{ir}^2}{[\sum_{i=1}^{n} |c_{ir}|]^2} \cdot \frac{1}{(1-q)^{s-1}}$$

Somit erhalten wir aus (14), (19), (23), (24)

$$(25)\quad P\{\sup_{\tau \varepsilon K_{(j)}} |W_{rn}(\tau,0)| \geq 2\varepsilon\} \leq A \, \frac{\sum_{i=1}^{n} c_{ir}^2}{[\sum_{i=1}^{n} |c_{ir}|]^2} \quad n^{\gamma}$$

mit

$$(26)\quad A = \frac{c}{b^2q(1-q)} + \frac{c}{9b^2q^2(1-q)^2}$$

Ferner gilt

$$(27)\quad \sup_{\tau \varepsilon K_{so}} |W_{rn}(\tau,0)| \leq \frac{\sum_{i=1}^{n} |c_{ir}|\,[u(x_i,0,c_i,d) + E\{u(x_i,0,c_i d)\}]}{\sqrt{n}}$$

mit $d = (1-q)^{so} \leq n^{-\gamma}$.

Also ist

$$(28)\quad P\{\sup_{\tau\varepsilon K_{s_o}} |W_{rn}(\tau,0)| \geq 2\varepsilon\} \leq$$

$$\leq P\{\sum_{i=1}^{n} |c_{ir}|\,[u(x_i,0,c_i,d)-E\{u(x_i,0,c_i,d)\}] \geq$$

$$\geq 2\varepsilon\sqrt{n} - 2\sum_{i=1}^{n} |c_{ir}|E\{u(x_i,0,c_i,d)\}\}$$

Da nach Voraussetzung $E\{u(x_i,0,c_i,d)\} \leq bd \leq bn^{-\gamma}$ ist, existiert ein n_1 so, dass für alle $n \geq n_1$

$$(29)\quad 2\sqrt{n}\varepsilon-2\sum_{i=1}^{n} |c_{ir}|E\{u(x_i,0,c_i,d)\} \geq 2\sqrt{n}\varepsilon-bn^{-\gamma}\sum_{i=1}^{n} |c_{ir}| \geq \sqrt{n}\varepsilon$$

Die Ungleichung von Tschebyscheff ergibt nun

$$(30)\quad P\{\sup_{\tau\varepsilon K_{s_o}} |W_{rn}(\tau,0)| \geq 2\varepsilon\} \leq \frac{c}{\varepsilon^2}\;\frac{1}{n}\sum_{i=1}^{n} c_{ir}^2\; n^{-\gamma}$$

Fassen wir (12), (15), (25) und (30) zusammen, so folgt für alle r, $1 \leq r \leq p$

$$(31)\quad P\{\sup_{\tau\varepsilon K_o} |W_{rn}(\tau,0)| \geq 2\varepsilon\} \leq 0(n^{-\gamma})+0\Bigg(\frac{\sum_{i=1}^{n} c_{ir}^2}{[\sum_{i=1}^{n} |c_{ir}|]^2}\; n^{\gamma}\log n\Bigg) =$$

$$= 0(n^{-\gamma})+0(n^{\gamma-1}\log n)$$

QED

Satz 5

Vor : (N1) bis (N5)

T_n genügt der Bedingung (1)

$P(|T_n| \leq \bar{d}_o) \to 1$

Beh :

(32) $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_i \psi(y_i - c_i' \theta_o) - \Lambda_n \sqrt{n}(T_n - \theta_o) \to 0$ in Wahrsch.

Bew : Sei wieder $\bar{d}_o = 1$, $\vartheta_o = 0$. Es gilt

(33) $Z_{ir}(T_n) = Z_{ir}(T_n) - Z_{ir}(0) - E\{Z_{ir}(T_n)\} + Z_{ir}(0) + E\{Z_{ir}(T_n)\}$

Es folgt mit gegen 1 strebender Wahrscheinlichkeit für $1 \leq r \leq p$

(34) $$\frac{\left|\sum_{i=1}^{n} [Z_{ir}(0) + E\{Z_{ir}(T_n)\}]\right|}{\sqrt{n} + \left|\frac{1}{n} \sum_{i=1}^{n} E\{Z_i(T_n)\}\right| \sum_{i=1}^{n} |c_{ir}|} \leq \sup_{|\tau| \leq 1} |W_n(\tau, 0)| + \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(T_n)\right|$$

Die rechte Seite strebt in Wahrscheinlichkeit gegen Null (Lemma 4 und Voraussetzung (1)), also auch die linke Seite.

Sei $0 < \varepsilon < \frac{1}{M^2} \leq 1$ vorgegeben und $K^2 := \frac{2}{\varepsilon} \sup_{F \varepsilon \mathfrak{F}} \int_{-\infty}^{+\infty} |\psi(x)|^2 F(dx)$

Für alle genügend grossen n und wegen der Ungleichung von Tschebyscheff sind die beiden Ungleichungen

$$(35) \qquad \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(0)\right| \leq K$$

und

$$(36) \qquad \left|\sum_{i=1}^{n} [Z_i(0)+E\{Z_i(T_n)\}]\right| \leq \varepsilon\left[\sqrt{n} + M\left|\sum_{i=1}^{n} E\{Z_i(T_n)\}\right|\right]$$

mit Wahrscheinlichkeiten $\leq \frac{\varepsilon}{2}$ verletzt, es gelten also beide gleichzeitig mit Wahrscheinlichkeiten $\geq 1 - \varepsilon$.

(36) impliziert

$$(37) \qquad \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\{Z_i(T_n)\}\right| \cdot (1-\varepsilon M) \leq \varepsilon + \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(0)\right|$$

Somit ist wegen (35)

$$(38) \qquad \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\{Z_i(T_n)\}\right| \leq \frac{K + \varepsilon}{1 - \varepsilon M}$$

Also gilt für alle genügend grossen n mit Wahrscheinlichkeit $\geq 1 - \varepsilon$

$$(39) \qquad \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Z_i(0) + E\{Z_i(T_n)\}]\right| \leq \varepsilon \, \frac{1 + MK}{1 - \varepsilon M}$$

Die rechte Seite kann beliebig klein gemacht werden, indem man ε klein genug wählt, das heisst

$$(40) \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Z_i(0) + E\{Z_i(T_n)\}] \rightarrow 0 \text{ in Wahrscheinlichkeit}$$

Durch Differentiation erhält man unmittelbar die Behauptung.

QED

Kontrollar :

Falls (32), (N3)(i),(N4) und (N5) erfüllt sind, gilt

(41) $\mathcal{L}(A_n^{-1}\ \Lambda_n\sqrt{n}\ (T_n-\theta_o)) \to \mathcal{N}(0,I_p)$, falls $n \to \infty$

Bew : Nach Eicker [5]

Sei $s(\mathfrak{F})$ die Menge aller Folgen von Fehlern mit Verteilungsfunktionen aus $\mathfrak{F}$: $X := \{x_i;\ i = 1,2,\ldots,\ \mathcal{L}(x_i)\varepsilon\mathfrak{F}\}\ \varepsilon s(\mathfrak{F})$ ist ein einzelnes Element dieser Menge.
Gleichmässig für alle Folgen $X\varepsilon s(\mathfrak{F})$ gelten :

(i) $\mathcal{L}[A_n^{-1}\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} c_i\psi(x_i)\}] \to \mathcal{N}(0,I_p)$, falls $n \to \infty$

(ii) Die einzelnen Summanden von $A_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} c_i\ (x_i)$ sind

gleichmässig klein dann und nur dann, wenn

(I) $\max\limits_{1\leqslant i\leqslant n} \frac{c_i'}{\sqrt{n}} D_n^{-1} \frac{c_i'}{\sqrt{n}} \to 0$, falls $n \to \infty$

(II) $\sup\limits_{F\varepsilon\mathfrak{F}} \int\limits_{|x|>t} \psi^2(x)F(dx) \to 0$, falls $t \to +\infty$

(III) $\inf\limits_{F\varepsilon\mathfrak{F}} \int_{-\infty}^{+\infty} \psi^2(x)F(dx) > 0$

(Die $\{c_{ir}\}$ werden in [5] nicht als beschränkt vorausgesetzt.)
Aus [5] folgt unmittelbar die Behauptung (41).

Würde man den zentralen Grenzwertsatz in seiner üblichen Form verwenden, so erhielte man Bedingungen, die gleichzeitig die Regressions-und die Fehlerfolgen enthielten. [4]

Das Korollar hat aber nur eine beschränkte praktische Bedeutung, da die Normierungsmatrix A_n^{-1} die Kenntnis der einzelnen Fehler voraussetzt.

Für die Methode der kleinsten Quadrate zeigte Eicher [5], dass durch Uebergang von den wahren Fehlern $x_i = y_i - c_i'\theta_o$ zu den scheinbaren Fehlern $e_i = y_i - c_i'T_n$ die asymptotische Normalität erhalten bleibt. Wesentlich beim Beweis ist ein Gesetz der grossen Zahlen für nicht-negative Zufallsgrössen. ([6], pag. 143)

Der Beweis lässt sich fast wörtlich auch auf andere genügend glatte Verlustfunktionen übertragen, wie z.B.

$$\rho_k(t) = \begin{cases} \frac{1}{2}\, t^2 & \text{für } |t| \leq k \\ k|t| - \frac{1}{2}\, k^2 & \text{für } t > k \end{cases}$$

indem man die Matrix $B_n^2 := \Lambda_n^{-1} A_n^2 \Lambda_n^{-1} = \Lambda_n^{-1} C_n' \sum_n C_n \Lambda_n^{-1}$

ersetzt durch $G_n^2 := L_n^{-1} C_n' S_n C_n L_n^{-1}$, wobei

$\psi(t) := \max(-k, \min(k,t))$, $S_n^2 := \mathrm{diag}(\psi^2(e_1), \psi^2(e_2),..,\psi^2(e_n))$

und

$$(L_n)_{rs} := \frac{1}{n} \sum_{i=1}^{n} c_{ir} c_{is} \psi'(y_i - c_i'T_n), \quad r,s = 1,2,\ldots,p$$

Literatur :

[1] Bessel, F.W. (1838). Untersuchungen über die Wahrscheinlichkeit von Beobachtungsfehlern. Astron. Nachr. Bd 15.

[2] Bickel, P.J. (1965). On some robust estimates of location. Ann. Math. Statist. 36.

[3] Czuber, E. (1891). Theorie der Beobachtungsfehler. Leipzig.

[4] Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. Ann. Math. Statist. 34 447 - 456.

[5] Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. Proc. Fith Berkeley Symp. on Math. Statist and Prob. Vol. 1 59 - 66.

[6] Gnedenko, B.V. and Kolmogorov A.N. (1954). Limit Distributions for Sums of Indipendent Random Variables. (Translated by K.L. Chung) Addison-Wesley, Cambridge.

[7] Huber, P.J. (1964). Robust estimation of a location parameter, Ann. Math. Statist. 35 73 - 101.

[8] Huber, P.J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. Proc. Fith. Berkeley Symp. on Math. Statist. and Prob. Vol. 1, 221-233.

[9] Laplace, P.S. (1812). Théorie analytique des probabilités. Paris.

[10] Linnik, Yu.V. (1961) Die Methode der kleinsten Quadrate in moderner Darstellung. Berlin

[11] Loève, M. (1967). Probability Theorie. (3nd ed.) Van Nostrand, Princeton

Lebenslauf

Ich wurde am 9. März 1942 in Willisau-Stadt geboren. Nach Schulen in Willisau und Luzern bestand ich 1961 an der Oberrealschule Luzern die Matura Typus C.

Anschliessend studierte ich an der ETH in Zürich Mathematik und erhielt im Frühjahr 1966 das Diplom und eine Stelle als Unterrichtsassistent von Prof. A. Huber. Ich betreute Studenten in den Analysisübungen und hatte während eines Semesters einen Lehrauftrag für eine Einführungsvorlesung in Wahrscheinlichkeitsrechnung und Statistik.

Gleichzeitig arbeitete ich bei Prof. P.J. Huber an einer Dissertation über robuste Schätzungen eines Regressionsparameters. Ein längerer Aufenthalt von Prof. P.J. Huber in den USA und mein Wegzug an die ETH Lausanne legten einen Wechsel in der Leitung meiner Dissertation nahe, die schliesslich auf Antrag von Prof. S. Chatterji von der EPFL entgegengenommen wurde.