ÜBER EINE ALLGEMEINE THEORIE DER ERWARTUNGSTREUEN SCHÄTZUNGEN

von

L. SCHMETTERER[1]

Es sei R eine Menge und S eine σ-Algebra von Teilmengen von R. $\mathfrak{P}$ sei eine nicht leere Klasse von Wahrscheinlichkeitsmaßen P, die über S definiert sind. Es sei g eine Abbildung von $\mathfrak{P}$ in den euklidischen R_1. Wir geben folgende

Definition 1. Eine S-meßbare Abbildung h von R in den R_1 soll erwartungstreu (oder auch erwartungstreue Schätzung) für g heißen, wenn der Erwartungswert von h für jedes P existiert und

$$E(h\,;P)^2 = g(P) \tag{1}$$

für alle $P \in \mathfrak{P}$.

Die Menge aller h, welche erwartungstreu für g sind, soll mit H_g bezeichnet werden.

Es gilt der

Satz 1. *H_g ist eine konvexe Menge.* —

Der Beweis ist trivial. Wenn H_g mehr als ein Element enthält, dann gibt es also unendlich viele erwartungstreue Schätzungen von g. Für den Statistiker erhebt sich dann das Problem der »optimalen Schätzung«, d. h. die Frage, welche erwartungstreue Schätzung aus H_g ausgewählt werden soll, um g »möglichst gut« zu schätzen. Diesem Problem sind die nachfolgenden Ausführungen gewidmet.

Es sei jedem $P \in \mathfrak{P}$ ein Banach-Raum B_P mit der Norm N_P zugeordnet. Wir nehmen nun an, daß die im folgenden betrachteten Durchschnittsmengen nicht leer sind und geben die

Definition 2. Ein Element $h_0 \in H_g \cap \bigcap_{P\in\mathfrak{P}} B_P$ soll gleichmäßig N_P-minimal heißen, wenn $N_P(h_0 - g(P)) \leqq N_P(h - g(P))$ für alle $h \in H_g \cap \bigcap_{P\in\mathfrak{P}} B_P$ und jedes $P \in \mathfrak{P}$.

Definition 3. Ein Element $h_0 \in H_g \cap B_{P_0}$ mit $P_0 \in \mathfrak{P}$ soll lokal N_{P_0}-minimal heißen, wenn $N_{P_0}(h_0 - g(P_0)) \leqq N_{P_0}(h - g(P_0))$ für alle $h \in H_g \cap B_{P_0}$.

Das wichtigste Beispiel für diese Definitionen ist der Fall, daß der Raum B_p für alle P mit dem Raum L_P^2 identisch ist, dessen Norm durch die Abbildung

[1] Hamburg.

[2] Es ist per definitionem $E(h: P) = \int_R h dP$.

$f \to (\int_R f^2\, dP)^{1/2}$ gegeben ist. In diesem Fall handelt es sich dann darum für alle P (oder für ein festes P_0 im lokalen Fall) die Streuung zu minimisieren. Für die Räume L^p_P, $p > 1$ habe ich dieses Problem kürzlich ausführlich untersucht [1]. Wir wollen nun zunächst die gleichmäßigen und lokalen Minimalschätzungen charakterisieren. Vorerst geben wir aber noch die folgende Erklärung :

Definition 4. V sei die Klasse aller S-meßbaren Abbildungen v von R in den R_1, deren Erwartungswerte für alle $P \in \mathfrak{P}$ existieren und welche die Gleichung erfüllen

$$E(v\,;P) = 0 \tag{2}$$

für alle $P \in \mathfrak{P}$.

Wir setzen nun voraus, daß B_P für jedes $P \in \mathfrak{P}$ glatt ist, d. h. daß durch jeden Punkt der Einheitssphäre genau eine Tangentialhyperebene hindurch geht. Dann gilt der

Satz 2. *Es seien alle B_P mit $P \in \mathfrak{P}$ glatt. $h_0 \in H_g \cap \bigcap_{P\in\mathfrak{P}} B_P$ ist genau dann gleichmäßig N_P-minimal, wenn das Gâteaux-Differential $L_P(h_0 - g(P)\,;\,v)$ der N_P-Norm von $h_0 - g(P)$ für alle Zuwächse $v \in V \cap \bigcap_{P\in\mathfrak{P}} B_P$ und jedes P verschwindet.* —

Zum Beweis bemerken wir zunächst, daß das Gâteaux-Differential der N_P-Norm für alle $P \in \mathfrak{P}$ existiert, da alle B_P als glatt vorausgesetzt sind.

Die Notwendigkeit der Bedingung ergibt sich unmittelbar aus der Definition der gleichmäßigen N_P-Minimalität und der Definition des Gâteaux-Differentiales.

Wir zeigen nun, daß die Bedingung auch hinreichend ist. Es sei also für ein $h_0 \in H_g \cap \bigcap_{P\in\mathfrak{P}} B_P$ $L_P(h_0 - g(P)\,;\,v) = 0$ für jedes $P \in \mathfrak{P}$ und für alle $v \in V \cap \bigcap_{P\in\mathfrak{P}} B_P$. Es sei weiter für ein $h \in H_g \cap \bigcap_{P\in\mathfrak{P}} B_P$ und ein $P_0 \in \mathfrak{P}$

$$N_{P_0}(h_0 - g(P_0)) > N_{P_0}(h - g(P_0))\,. \tag{3}$$

Nun ist $h - h_0 \in V \cap \bigcap_{P\in\mathfrak{P}} B_P$ und voraussetzungsgemäß gilt für $t \to 0$ $N_{P_0}(h_0 - g(P_0) + t(h - h_0)) - N_{P_0}(h_0 - g(P_0)) = o(t)$. Es sei nun $t > 0$. Es folgt für jedes $\varepsilon > 0$ und hinreichend kleines $t > 0$

$$-\varepsilon t \leqq N_{P_0}((1-t)(h_0 - g(P_0) + t(h - g(P_0)) - N_{P_0}(h_0 - g(P_0)) \leqq \varepsilon t\,. \tag{4}$$

Benützen wir die linke Seite von (4), dann folgt

$$\begin{aligned} -\varepsilon t \leqq (1-t)\, N_{P_0}(h_0 - g(P_0)) + t\, N_{P_0}(h - g(P_0)) - N_{P_0}(h_0 - g(P_0)) &= \\ = t\left[N_{P_0}(h - g(P_0)) - N_{P_0}(h_0 - g(P_0))\right]. \end{aligned}$$

Die daraus hergeleitete Ungleichung $-\varepsilon \leqq N_{P_0}(h - g(P_0)) - N_{P_0}(h_0 - g(P_0))$ steht aber im Widerspruch zu (3).

Selbstverständlich gilt ein entsprechender Satz auch für lokal N_{P_0}-minimale Schätzungen, auf dessen Formulierung wir verzichten. Hinsichtlich der Eindeutigkeit von Minimalschätzungen gilt der

Satz 3. *Wenn B_P für jedes $P \in \mathfrak{P}$ strikt konvex ist, dann gibt es (bis auf $\mathfrak{P}$-Nullmengen*[3] *höchstens eine gleichmäßig N_P-minimale Schätzung für gegebenes g.* —

Den einfachen Beweis habe ich in einer früheren Arbeit gegeben [2]. Wieder gilt ein analoger Satz für den Fall lokal minimaler Schätzungen.

Wir beschränken uns nun auf den Fall, daß ein für $\mathfrak{P}$ dominantes Maß μ existiert und $\mu \in \mathfrak{P}$ ist. Wir bezeichnen die nach dem Satz von Radon-Nikodym existierenden Dichten von P bezüglich μ mit f_P und nehmen stets an, daß f_P für jedes P dem zu B_μ konjugierten Raum B_μ^* angehört. Mit anderen Worten, wir setzen voraus, daß durch die Abbildung $k \to \int_R k f_P \, d\mu$ für $k \in B_\mu$ ein lineares und stetiges Funktional über B_μ definiert wird. Wir beschäftigen uns nun mit lokal N_μ-minimalen Schätzungen und beweisen zunächst das

Lemma 1. *Unter den angeführten Voraussetzungen ist die Menge $V_\mu = V \cap B_\mu$ eine abgeschlossene lineare Mannigfaltigkeit in B_μ.*

Es ist trivial, daß V_μ ein Vektorraum ist. Die Abgeschlossenheit von V_μ folgt ebenfalls ganz leicht: Es sei $v_n \in V_\mu$ für $n \geqq 1$ und $N_\mu(v_n - v) \to 0$. Es folgt unter Benützung von (2)

$$\left|\int_R v f_P \, d\mu\right| = \left|\int_R v_n f_P \, d\mu + \int_R (v - v_n) f_P \, d\mu\right| \leqq N_\mu(v_n - v) \, N_\mu^*(f_P) \to 0$$

für jedes $P \in \mathfrak{P}$. Dabei bedeutet N_μ^* natürlich die Norm in B_μ^*. Wir betrachten nun die Menge G_μ aller Abbildungen g_k von $\mathfrak{P}$ in den R_1, welche durch

$$P \to \int_R k f_P \, d\mu$$

mit $k \in B_\mu$ gegeben sind.

Lemma 1 hat zur Folge, daß der Quotientenraum $Q_\mu = B_\mu / V_\mu$ wieder ein Banachraum ist, wenn man die Norm N_{Q_μ} von Q_μ durch

$$N_{Q_\mu}(y) = \inf_{y = \varphi(x)} N_\mu(x) \tag{5}$$

definiert. φ ist die kanonische Abbildung von B_μ in Q_μ. Aus der Definition von V_μ folgt sofort, das

Lemma 2. *G_μ und Q_μ können in natürlicher Weise eineindeutig aufeinander bezogen werden.*

Nunmehr ist es möglich, das folgende Resultat zu beweisen, welches ein Ergebnis von Stein [3] und dessen Weiterführung in der erwähnten Arbeit [1] verallgemeinert:

Satz 4. *Die Klasse $\mathfrak{P}$ werde durch ein Maß μ dominiert, das selbst zu $\mathfrak{P}$ gehöre, f_P, die Dichte von P bezgl. μ, gehöre für jedes $P \in \mathfrak{P}$ zu B_μ. B_μ sei ein glatter Banach-Raum. $h \in H_g \cap B_\mu$ ist genau dann lokal N_μ-minimal, wenn eine Abbildung T von G_μ in den R_1 existiert mit der Eigenschaft daß*

$$T(g_k) = L_\mu(h - g(\mu)\,; k)$$

für alle $k \in B_\mu$, wobei L_μ natürlich das Gâteaux-Differential der Norm N_μ bezeichnet.

[3] Darunter verstehen wir eine Menge $A \in S$, für die $P(A) = 0$ für alle $P \in \mathfrak{P}$ gilt.

Der Beweis beruht auf Lemma 1, Lemma 2 und dem folgenden

Lemma 3. *B sei ein Banach-Raum, M eine abgeschlossene lineare Mannigfaltigkeit von B und B^* der zu B konjugierte Raum. Es sei weiter $M^0 \subset B^*$ der Annihilator von M. Q sei der Quotienten-Raum von B nach M und Q^* der zu Q konjugierte Raum. Es sei φ die kanonische Abbildung von B in Q. Dann ist die zu φ adjungierte Abbildung φ^* eine eineindeutige, lineare und isometrische Abbildung von Q^* auf M^0* [4].

Nun ist es sehr bekannt, daß die Abbildung $k \to L_\mu(h - g(\mu)\,;\,k)$ linear ist, und es ist leicht zu zeigen, daß sie auch beschränkt ist. Jetzt kann aber der Beweis des Satzes 4 wortwörtlich so geführt werden, wie ich das für den Fall, daß B_μ der Raum $L^p_\mu\,(p > 1)$ ist, in [1] gezeigt habe. Dort finden sich auch Beispiele, wie man diesen Satz zur Konstruktion von lokal minimalen Schätzungen benützen kann.

Wir wollen uns nun der Frage der Existenz lokal minimaler Schätzungen zuwenden. Hierzu machen wir wieder die Voraussetzung, daß ein für die Klasse $\mathfrak{P}$ dominantes Maß μ existiert, das zu $\mathfrak{P}$ gehört. Überdies nehmen wir an, daß V_μ eine abgeschlossene lineare Mannigfaltigkeit ist. Lemma 2 und (5) lehren dann, daß das Problem der Existenz einer lokal N_μ-minimalen Schätzung äquivalent ist mit der Frage, ob das Infimum in (5) angenommen wird. Es gilt nun der

Satz 5. *Wenn die angegebenen Voraussetzungen erfüllt sind und V_μ ein reflexiver Banach-Raum ist, dann existiert für alle $g_k \in G_\mu$ eine lokal N_μ-minimale erwartungstreue Schätzung.*

Zum Beweise sei k ein beliebiges Element aus B_μ. Mit $y_k \in Q_\mu$ bezeichnen wir die Klasse $\{k + v\}$, $v \in V_\mu$. Es sei l_n eine Folge von Elementen aus $\{k + v\}$ mit $N_\mu(l_n) \to N_{Q_\mu}(y_k)$. Es ist leicht einzusehen, daß mit V_μ auch der Banachraum $B_\mu(k)$ der durch k und V_μ erzeugt wird (und die durch N_μ induzierte Norm besitzen soll), reflexiv ist. Somit existiert nach einem bekannten Satz (vgl. z. B. [5]) ein $l \in B_\mu(k)$, so daß l_n schwach gegen l konvergiert. Die Menge der $\{k + v\}$ ist aber konvex und abgeschlossen in $B_\mu(k)$. Somit gehört l zu $\{k + v\}$ (vgl. [6]) und $N_\mu(l) = N_{Q_\mu}(y_k)$ und das war zu beweisen.

Da jede abgeschlossene Mannigfaltigkeit eines reflexiven Banach-Raumes selbst reflexiv ist, folgt aus Satz 5 sofort das

Corollar.[4] *Wenn B selbst reflexiv ist, dann existiert unter den angegebenen Voraussetzungen für alle $g_k \in G$ eine lokal N_μ-minimale Schätzung.*

Für den Fall, daß $B_\mu = L^p_\mu$, $p > 1$, wurde dieses Corollar auf anderem Wege von BARANKIN [8] bewiesen.

Wir wollen nun zum Schluß noch eine Anwendung geben. BRUDNO [9] hat vor kurzem mittels einer komplizierten Beweismethode folgende Verallgemeinerung eines Satzes von Markov gezeigt: Es seien $y_1, \ldots, y_m$ unabhängige zufällige Variable, die beziehungsweise nach einer Normalverteilung $N\left(\sum_{i=1}^{n} \alpha_i\, x_{ik}, \sigma_k^2\right)$, $1 \leq k \leq m$, verteilt sind. Es sei $m > n$, die x_{ik}, $1 \leq i \leq n$, $1 \leq k \leq m$ seien gegebene reelle Zahlen, deren Matrix $X = (x_{ik}/\sigma_k)$ den Rang n habe und die α_i seien Parameter, die für $1 \leq i \leq n$ der Bedingung

[4] Nach Fertigstellung dieser Arbeit hat mich Herr Dr. BAUER, Hamburg, darauf aufmerksam gemacht, daß in etwas anderem Zusammenhang ein solcher Satz schon früher gegeben wurde (vgl. [7]).

$-\infty < \alpha_i < \infty$ genügen. Es seien reelle Zahlen c_i, $1 \leq i \leq n$, gegeben. Dann gilt : In der Klasse der über dem R_m definierten erwartungstreuen Schätzungen für $\sum_{i=1}^{n} c_i \alpha_i$ ist die durch die Gaußsche Methode der kleinsten Quadrate gegebene Schätzung gleichmäßig $L^2_{P_\alpha}$-minimal. Hierbei sei P_α für $\alpha = (\alpha_1, \ldots, \alpha_n)$ die Verteilung, deren Dichte durch

$$\prod_{k=1}^{m} \left(\frac{1}{\sqrt{2\pi\sigma_k}}\right) \exp\left[-\frac{1}{2\,\sigma_k}\left(y_k - \sum_{i=1}^{n} \alpha_i x_{ik}\right)^2\right] \text{für jedes } y = (y_1, \ldots, y_m) \in R_m$$

gegeben ist und $\mathfrak{P}$ die Menge aller P_α (bei festen $\sigma_1, \ldots, \sigma_m$).

Dieser Satz folgt aber sofort aus Satz 2 und 3. Man betrachte nämlich die Menge V aller v, welche für alle $P_\alpha \in \mathfrak{P}$ der Gleichung

$$\int_R v\,dP_\alpha = 0 \tag{6}$$

genügen. Für alle α, die etwa einem beliebigen offenen Quader des R_n angehören, und alle $v \in V \cap \bigcap_\alpha L^2_{P_\alpha}$ folgt aber aus (6) durch Differentiation nach α_j sofort

$$\int_{R_m} v \left[\sum_{k=1}^{m} \frac{1}{\sigma_k}\left(y_k - \sum_{i=1}^{n} \alpha_i x_{ik}\right) x_{jk}\right] dP_\alpha(y) = 0, \qquad 1 \leq j \leq n.$$

Daraus folgt aber wieder aus (6) für $1 \leq j \leq n$

$\int_{R_m} v \sum_{k=1}^{m} y_k x_{jk}/\sigma_k \, dP_\alpha(y) = 0$ oder auch, wenn wir mit X' die zu X transponierte Matrix bezeichnen und $y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ schreiben:

$$\int_{R_m} vX'y\,dP_\alpha(y) = 0 \quad \text{oder auch} \quad \int_{R_m} v(X'X)^{-1} X'y\,dP_\alpha(y) = 0.$$[5]

Nach Multiplikation mit $c' = (c_1, \ldots, c_n)$ erhalten wir also

$$\int_{R_m} vc'(X'X)^{-1} X'y\,dP_\alpha(y) = 0$$

für alle $P_\alpha \in \mathfrak{P}$ und alle $v \in V \cap \bigcap_\alpha L^2_{P_\alpha}$. Nach Satz 3 ist daher $c'(X'X)^{-1} X'y$ gleichmäßig $L^2_{P_\alpha}$-minimale Schätzung für $\sum_{i=1}^{n} c_i \alpha_i$ und nach Satz 2 auch die einzige. Bekanntlich ist aber durch $(X'X)^{-1} X'y$ der Lösungsvektor der Gaußschen Gleichung gegeben und somit ist der Satz bewiesen.

(Eingegangen: 11. Januar, 1961.)

[5] Unter dem Integral einer Matrix verstehen wir natürlich die Matrix der Integrale.

LITERATURVERZEICHNIS

[1] SCHMETTERER, L.: „On unbiased estimation". *Ann. Math. Stat.* **31** (1960) 1154—1163.
[2] SCHMETTERER, L.: „Bemerkungen zur Theorie der erwartungstreuen Schätzungen". *Mitteil. -Bl. math. Statistik* **9** (1957) 147—152.
[3] STEIN, C.: „Unbiased estimates with minimum variances". *Ann. Math. Stat.* **21** (1950) 406—415.
[4] BOURBAKI, N.: *Livre V, Espaces Vectoriels Topologiques,* Chapitre III—V, 1955, 115.
[5] HILLE, E. und PHILLIPS, R. S.: American Math. Soc. Colloquium Publ. XXXI, *Functional Analysis and semi-groups.* Revised Edition, 1957, 37.
[6] L. c. [5], 36.
[7] HIRSCHFELD, R. A.: „On best approximations in normed vector spaces". *Nieuw Archief voor Wiskunde* (3) **6** (1958) 41—51.
[8] BARANKIN, E. W.: "Locally best unbiased estimates". *Ann. Math. Stat.* **20** (1949) 477—502.
[9] БРУДНО, А. Л.: "К дисперсионному обоснованию метода наименьших квадратов." *Mat. Sbornik* **43** (**85**) (1957) 37—48.

ОБ ОДНОЙ ОБЩЕЙ ТЕОРИИ НЕСМЕЩЕННЫХ ОЦЕНОК

L. SCHMETTERER

Резюме

В настоящей статье дополняется теория прежней работы опубликованной автором в журнале *The Annals of Math. Stat.* **31** (1960) 1154-1163. Пусть $\mathfrak{P}$ непустой класс мер вероятностей и g действительная функция определенная в множестве $\mathfrak{P}$. Пусть дальше H_g класс несмещенных оценок для g. Допустим, что каждой мере P, принадлежащей множеству $\mathfrak{P}$, соответсвует пространство Банаха B_P с нормой N_P. Элемент $h_0 \in H_g \cap \bigcap_{P \in \mathfrak{P}} B_P$ называется равномерно N_P -минимальной, если $N_P(h_0 - g(P)) \leqq N_P(h - g(P))$ при всех элементах $h \in H_g \cap \bigcap_{P \in \mathfrak{P}} B_P$ и каждой мере $P \in \mathfrak{P}$. $h_0 \in H_g \cap B_{P_0}$ называется локально N_{P_0}-минимальной, если $N_{P_0}(h_0 - g(P_0)) \leqq N_{P_0}(h - g(P_0))$ при всех элементах $h \in H_g \cap B_{P_0}$ и мере $P_0 \in \mathfrak{P}$. В этой статье даются необходимое и достаточное условия для существования и однозначности равномерно и локально минимальных несмещенных оценок. Известные теоремы БАРАНКИНА, ЛЕМАННА, ШТЕИНА и других можно рассматривать как частные случаи этих общих результатов. В качестве примера дается краткое доказательство теоремы БРУДНО о методе наименьших квадратов.