Werk

Titel: Über nichtparametrische Methoden in der Mathematischen Statistik.

Autor: Schmetterer, Leopold

Jahr: 1959

PURL: https://resolver.sub.uni-goettingen.de/purl?37721857X_0061|log15

Kontakt/Contact

Digizeitschriften e.V.
SUB Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen

info@digizeitschriften.de

Schmetterer, Leopold
Jahresbericht d. DMV
Bd. 61 (1959) S. 104—126

Über nichtparametrische Methoden in der Mathematischen Statistik.*)

Von Leopold SCHMETTERER in Hamburg.

Die sogenannten nichtparametrischen Methoden in der Mathematischen Statistik gehören im allgemeinen der neueren Zeit an, doch reichen manche Ansätze weit zurück. Eine Reihe moderner nichtparametrischer Tests beginnen sich neben den klassischen Methoden einen festen Platz in den praktischen Anwendungen zu erobern. Hier soll aber vor allem von den allgemeineren Gesichtspunkten die Rede sein, deren Herausarbeitung meist neueren Datums und keineswegs abgeschlossen ist. Nach keiner Richtung hin kann in diesem Bericht Vollständigkeit angestrebt werden. Wir sehen z. B. von der Theorie der Konfidenzbereiche ab, betrachten auch keine speziellen Mehrentscheidungsprobleme und gehen auch nicht auf asymptotische [0] Fragen ein. Wir beschränken uns vielmehr im wesentlichen auf zwei Aspekte: Die Theorie der erwartungstreuen Schätzungen und die Theorie der nichtparametrischen Tests.

I. Man kann selbstverständlich die mathematische Theorie der Statistik entwickeln ohne auf die Anwendungen Bezug zu nehmen. Diesen Vorgang möchte ich auch hier einhalten. Da aber gewisse mathematische Fragestellungen in der Statistik erst vom Blickpunkt ihrer Anwendung her verständlich werden, soll kurz ein einfaches und typisches Problem der Mathematischen Statistik geschildert werden: Es seien n Meßwerte $x_1, \ldots x_n$ gegeben, die wir als geordnetes n-Tupel reeller Zahlen oder Punkt des R_n auffassen können und die wir als Stichprobe bezeichnen. Wir machen die Annahme, daß die Meßwerte ausschließlich „zufälligen Fehlern" unterworfen sind. Über die Natur des hierbei waltenden Fehlergesetzes möchten wir an Hand der Stichprobe möglichst verläßlichen Aufschluß gewinnen. Hierzu bedienen wir uns einer Modellvorstellung: $x_1, \ldots, x_n$ fassen wir als Realisationen unabhängiger zufälliger Variabler [1] auf, die alle nach demselben, jedoch unbekannten Verteilungsgesetz verteilt sind, eben dem Fehlergesetz. Weiter nehmen wir an, daß das unbekannte Verteilungsgesetz, welches wir

*) Dieser Bericht stellt im Wesentlichen die Wiedergabe eines Referates dar, welches ich bei der Tagung der DMV am 9. 9. 1957 in Dresden gehalten habe.

auch als richtiges Verteilungsgesetz bezeichnen, einer wohldefinierten Klasse K von Verteilungsgesetzen angehört, wobei wir stets den trivialen Fall ausschließen, daß K nur ein Element enthält. Es soll nun, gegründet auf die vorliegende Stichprobe, eine Entscheidung darüber getroffen werden, welches Fehlergesetz aus K das richtige ist. Die Konsequenzen einer solchen Entscheidung werden einerseits eine Funktion der Stichprobe sein, anderseits aber auch vom richtigen Fehlergesetz abhängen. Selbstverständlich wird man wünschen, die in irgendeinem Sinne beste Entscheidung zu wählen. Die späteren Definitionen und Überlegungen dienen fast alle diesem Ziel.

Die eben erwähnte Problemstellung tritt vielfach in einer etwas modifizierten Form auf. Man betrachtet eine nicht leere Teilklasse $k \subset K$, die man als Nullhypothese bezeichnet. Die Aufgabe besteht nun darin, auf Grund der gegebenen Stichprobe zu entscheiden, ob das richtige Verteilungsgesetz zu k gehört oder nicht. Man nennt diese Entscheidung das Testen der Nullhypothese k. Wieder wird man sich für die Konsequenzen einer solchen Entscheidung interessieren.

Einige Schwierigkeiten bereitet es, den Begriff „nichtparametrisch" so zu umschreiben, daß er allen (nicht scharf umrissenen) Vorstellungen entspricht, die man mit ihm üblicherweise verbindet. Die klassischen Methoden der Mathematischen Statistik sind vor allen Dingen für solche Klassen K von Verteilungsgesetzen entwickelt worden, die sich in natürlicher Weise topologisch in endlich-dimensionale euklidische Räume abbilden lassen [2]. Handelt es sich genau um einen m-dimensionalen Raum, dann sagt man, es liege ein Problem mit m Parametern vor. Wir kennzeichnen nun die *nichtparametrischen Methoden* negativ dadurch, daß wir von einer derartigen Voraussetzung keinen Gebrauch machen. Wir weichen damit allerdings vom vielfach üblichen, jedoch meist vagen Sprachgebrauch etwas ab.

Wir gehen nun zu einer präzisen Fassung der andeutungsweise entwickelten Begriffe über. Wir führen folgende Bezeichnungen ein, die wir später ohne Erklärung benützen: Mit $(R, \mathfrak{B})$ bezeichnen wir einen *meßbaren Raum*, d. h. R ist eine beliebige Menge und $\mathfrak{B}$ ein σ-Körper von Teilmengen aus R, der stets R selbst enthalten soll. Die Mengen $B \in \mathfrak{B}$ bezeichnen wir als meßbar. Über $\mathfrak{B}$ sei eine (nicht leere) Klasse K von Wahrscheinlichkeitsmaßen P definiert. Mit $W(B; P)$ bezeichnen wir die Wahrscheinlichkeit der Menge B, wenn das Wahrscheinlichkeitsmaß $P \in K$ zugrundeliegt. Die Elemente von R bezeichnen wir mit x. Wenn $u(x)$ eine zufällige Variable ist, dann benützen wir für den Erwartungswert bei Zugrundelegung des Wahrscheinlichkeitsmaßes $P \in K$ die Schreibweise $E(u(x); P)$. Dabei ist $E(u(x); P) = \int_R u(x)\, dP$.

Die Gesamtheit aller geordneten n-Tupel reeller Zahlen $(x_1, \ldots, x_n)$ bezeichnen wir mit R_n. In der Bezeichnung $(R_n, \mathfrak{B})$ bedeutet $\mathfrak{B}$ stets den σ-Körper der euklidischen Borelschen Mengen.

Sei M eine beliebige Menge. Das n-fache cartesische Produkt der Menge M mit sich selbst bezeichnen wir stets mit $M^{(n)}$. Wenn $(M, \mathfrak{S})$ ein meßbarer Raum ist, dann bezeichnen wir den durch $(M, \mathfrak{S})$ induzierten Produktmaßraum mit $(M^{(n)}, \mathfrak{S}^{(n)})$ [3]. Wenn über $\mathfrak{S}$ eine Klasse K von Wahrscheinlichkeitsmaßen P definiert ist, dann bezeichnen wir die Menge aller Produktmaße $P \times P \times \cdots \times P$, $(P \in K)$, die über $\mathfrak{S}^{(n)}$ definiert sind, mit $K^{(n)}$. Von diesen Bezeichnungen weichen wir nur insofern ab, als wir das cartesische Produkt des R_1 mit sich selbst mit R_n bezeichnen, wie wir bereits oben gesagt haben.

Für alle Borelschen Mengen des R_n ist ein Wahrscheinlichkeitsmaß eindeutig definiert, wenn man die zugehörige Verteilungsfunktion $F(x_1, \ldots, x_n)$ kennt. Die *Gesamtheit aller eindimensionalen Verteilungsfunktionen* $F(x)$ bezeichnen wir mit $\mathfrak{F}$. Die *Gesamtheit aller stetigen eindimensionalen Verteilungsfunktionen* bezeichnen wir mit S. Die Teilmengen aus $\mathfrak{F}$ bzw. S, welche durch die Bedingung

$$F(0) = p, \quad 0 < p < 1 \tag{1}$$

gekennzeichnet sind, bezeichnen wir mit ${}_p\mathfrak{F}$ bzw. mit ${}_pS$. Die Teilmengen aus $\mathfrak{F}$ bzw. S, welche durch

$$F(x) = 1 - F(-x-0) \qquad -\infty < x < +\infty \tag{2}$$

gekennzeichnet sind, bezeichnen wir mit ${}_s\mathfrak{F}$ bzw. ${}_sS$.

Die Gesamtheit der Produktmaße, welche durch die Verteilungsfunktionen der Gestalt $\prod_{i=1}^{n} F_i(x_i)$ mit $F_i(x) \in \mathfrak{F}$ definiert sind, bezeichnen wir mit $\mathfrak{F}^{(n)}_{1,\ldots,n}$. $\mathfrak{F}^{(n)}$ bedeutet ähnlich die Gesamtheit aller durch $\prod_{i=1}^{n} F(x_i)$ definierten Produktmaße mit $F(x) \in \mathfrak{F}$. Analog sind ${}_p\mathfrak{F}^{(n)}_{1,\ldots,n}$ und andere zu verstehen. Weiter bezeichnen wir mit V_a die Menge aller absolut stetigen und mit V_d die Menge aller diskreten eindimensionalen Wahrscheinlichkeitsmaße. Die Bedeutung von $V_a^{(n)}$ oder $V_d^{(n)}$ versteht sich von selbst.

Wir gehen nun von einem meßbaren Raum $(R, \mathfrak{B})$ aus und bezeichnen in diesem Zusammenhang R auch als Stichprobenraum. C sei eine Menge, die wir als *Raum der Entscheidungen* c bezeichnen. Wir betrachten die Menge aller Abbildungen $\gamma(x)$ (oder auch eine geeignete Teilmenge derselben) von R in C. Jede solche Abbildung $\gamma(x)$ soll eine *Entscheidungsfunktion* [4] genannt werden. Weiter sei eine über $C \times K$

definierte Funktion $G(c, P)$ gegeben, welche wir als Gewichtsfunktion bezeichnen. Wir nehmen nun an, daß $E(G(\gamma(x), P); P)$ (mindestens bei Beschränkung auf eine geeignete Teilmenge von Entscheidungsfunktionen) stets für alle $P \in K$ existiert. Offenbar sind implizit in dieser Forderung Meßbarkeitsvoraussetzungen für die Funktion $G(\gamma(x), P)$ für jedes feste $P \in K$ enthalten. Für diesen Erwartungswert, der offenbar von der gewählten Entscheidungsfunktion $\gamma(x)$ und dem Wahrscheinlichkeitsmaß P abhängt, schreiben wir $R(\gamma, P)$ und bezeichnen ihn als *Risikofunktion*.

Besonders für die Testtheorie ist es wichtig, dieses Konzept zu erweitern. Es sei $(C, \mathfrak{C})$ ein meßbarer Raum. Es ist zweckmäßig anzunehmen, daß für alle Entscheidungen $c \in C$ die Mengen $\{c\}$ zu $\mathfrak{C}$ gehören. Über $\mathfrak{C}$ sei eine Menge E von Wahrscheinlichkeitsmaßen e gegeben. Wir nehmen an, daß E zu jedem $c \in C$ ein Wahrscheinlichkeitsmaß e_c enthält, welches durch $e_c(\{c\}) = 1$ definiert ist. Als (verallgemeinerte) Entscheidungsfunktion bezeichnen wir jede Abbildung $\varepsilon(x)$ von R in E, wofür wir auch $\varepsilon(c \mid x)$ schreiben, wenn wir daran erinnern wollen, daß $\varepsilon(x)$ für jedes x ein Wahrscheinlichkeitsmaß über $\mathfrak{C}$ liefert. Identifiziert man das Wahrscheinlichkeitsmaß e_c mit c, dann erkennt man, daß in der Menge aller verallgemeinerten Entscheidungsfunktionen die vorhin definierten Entscheidungsfunktionen als Teilmenge enthalten sind. Wir werden daher künftighin nur mehr von Entscheidungsfunktionen schlechthin sprechen. Die Risikofunktion $R(\varepsilon, P)$ ist dann durch $E(L(\varepsilon(x), P); P)$ gegeben mit

$$L(\varepsilon(x), P) = \int_C G(c, P)\, d\varepsilon(c \mid x).$$

Wir geben nun eine Reihe wichtiger Definitionen, welche den Begriff einer „guten Entscheidungsfunktion" präzise fassen sollen. $\varepsilon(x)$ heißt *nicht schlechter* als $\varepsilon_1(x)$ bezüglich K, wenn

$$R(\varepsilon, P) \leq R(\varepsilon_1, P) \quad \text{für alle } P \in K \text{ gilt.}$$

$\varepsilon(x)$ heißt *besser* als $\varepsilon_1(x)$, wenn in dieser Ungleichung mindestens für ein P das $<$-Zeichen gilt.

Eine Klasse Q von Entscheidungsfunktionen heißt *vollständig*, wenn es zu jeder nicht in Q liegenden Entscheidungsfunktion eine bessere aus Q gibt.

Eine Entscheidungsfunktion $\varepsilon(x)$ heißt *minimal*, wenn sie nicht schlechter bezüglich K ist als jede andere Entscheidungsfunktion, d. h. wenn

$$R(\varepsilon, P) \leq R(\varepsilon_1, P)$$

für alle $P \in K$ und alle ε_1 gilt.

Eine Entscheidungsfunktion $\varepsilon(x)$ heißt *minimax*, wenn

$$\sup_{P \in K} R(\varepsilon, P) \leq \sup_{P \in K} R(\varepsilon_1, P)$$

für alle Entscheidungsfunktionen ε_1 gilt.

Alle diese Definitionen sind sinngemäß auf den Fall zu übertragen, daß man sich auf eine Teilmenge von Entscheidungsfunktionen beschränkt.

Eine Entscheidungsfunktion $\varepsilon(x)$ heißt *unverfälscht* [5], wenn

$$E(L(\varepsilon(x), P'); P) \geq E(L(\varepsilon(x), P); P) \tag{3}$$

für alle $P, P' \in K$ gilt.

Wir wollen nun den Begriff der *Invarianz* [6] einführen. $\mathfrak{G}$ sei eine beliebige Gruppe. γ_R sei ein Homomorphismus von $\mathfrak{G}$ in die Menge aller eineindeutigen meßbaren Transformationen des meßbaren Raumes $(R, \mathfrak{B})$ in sich. γ_C sei ein entsprechender Homomorphismus für den meßbaren Raum $(C, \mathfrak{C})$ und schließlich sei γ_K ein Homomorphismus von $\mathfrak{G}$ in die Menge aller eineindeutigen Transformationen von K in sich. Für jedes $g \in \mathfrak{G}$ bezeichnen wir $\gamma_R(g)$ mit g_R usw. Die Gruppe $\mathfrak{G}$ heißt für ein Entscheidungsproblem zulässig, wenn folgendes gilt:

1. $g_K P(g_R B) = P(B)$ für alle $B \in \mathfrak{B}$, alle $P \in K$ und alle $g \in \mathfrak{G}$.
2. $G(g_C c, g_K P) = G(c, P)$ für alle $c \in C$, $P \in \mathrm{X}$ und $g \in \mathfrak{G}$.

Wir definieren nun für jedes $g \in \mathfrak{G}$ eindeutige Transformationen g_E von E in sich gemäß

$$g_E \varepsilon(c \mid x) = \varepsilon(g_C c \mid g_R x).$$

Es ist, wie man unmittelbar erkennt,

$$R(g_E \varepsilon, P) = R(\varepsilon, g_K P).$$

Eine Entscheidungsfunktion heißt *invariant*, wenn $g_E \varepsilon(c \mid x) = \varepsilon(c \mid x)$ für alle $x \in R$. Diesen Begriff wollen wir später noch etwas verallgemeinern.

II. Sei $m(P)$ ein beliebiges über K definiertes Funktional. Wir suchen eine über R definierte reellwertige meßbare Funktion $h(x)$, für die $E(h(x); P)$ für alle $P \in K$ existiert und

$$E(h(x); P) = m(P) \tag{4}$$

für alle $P \in K$ ist. Jede solche Funktion $h(x)$ nennen wir *erwartungstreue Schätzung* für $m(P)$ [7]. Auf die naheliegende Verallgemeinerung dieses Begriffes auf vektorwertige Funktionale mit Werten aus einem R_n kommen wir später kurz zu sprechen. Aus (4) folgt bekanntlich:

$y_i = h_i(x)$ $(i = 1, 2, \ldots)$, $x \in R$ seien unabhängige zufällige Variable mit derselben Verteilung wie $h(x)$. Dann strebt für $l \to \infty$ $(y_1 + \cdots + y_l)/l$ für jedes feste $P \in K$ der Wahrscheinlichkeit nach gegen $m(P)$ [8].

Wir können also in der Terminologie der Mathematischen Statistik sagen: $(h_1(x) + \cdots + h_l(x))/l$ ist eine *konsistente Schätzfunktionenfolge* für $m(P)$ über K und darin liegt ein Gutteil der praktischen Bedeutung von (4).

Ist nun x eine Stichprobe mit unbekannter Verteilung P aus K, dann ist es naheliegend, $m(P)$ für dieses P durch $h(x)$ zu ersetzen. Im allgemeinen besitzt aber (4) mehrere verschiedene Lösungen. Es erhebt sich die Frage, welche Lösung man auswählen soll, um $m(P)$ „möglichst gut" durch $h(x)$ zu ersetzen.

Es sei H die Klasse aller Lösungen von (4), die im folgenden stets als nichtleer vorausgesetzt wird. Wir definieren im Sinne von I. eine Gewichtsfunktion $G(z, P)$, die über $R_1 \times K$ definiert ist und betrachten die zugehörige Risikofunktion

$$R(h, P) = E(G(h(x), P); P)$$

für alle $P \in K$ und $h(x) \in H$, wobei wir stets der Einfachheit halber annehmen, daß $R(h, P)$ für alle $h(x) \in H$ existiert. Wir fassen also jedes $h(x) \in H$ als Entscheidungsfunktion im Sinne von I. auf und fragen nach den minimalen $h(x) \in H$, die wir auch als *erwartungstreue Minimalschätzungen* bezeichnen.

Unter Beachtung der trivialen Tatsache, daß H konvex ist, folgt leicht ein Satz, der in einfachster Form bei Lehmann und Scheffé steht [9]:

Satz 1: Sei $G_1(z)$ eine über dem R_1 definierte nicht negative und strikt konvexe Funktion. Sei $G(h, P) = G_1(h - m(P))$, dann existiert höchstens eine Minimalschätzung $h(x) \in H$.

Selbstverständlich verstehen sich hier und im folgenden alle Eindeutigkeitsaussagen nur bis auf Nullmengen.

Es sei $r \geqq 1$ und N^r die Klasse der erwartungstreuen Schätzungen $v(x)$ für das über K definierte identisch verschwindende Funktional mit $E(|v|^r; P) < \infty$ für alle $P \in K$. Wir betrachten den Raum $L^r(P)$, der bezüglich P zur r-ten Potenz integrierbaren reellwertigen Funktionen mit der üblichen Norm. Dann gilt der

Satz 2: Es sei $r > 1$ und $H \subset L^r(P)$ für alle $P \in K$ und die Gewichtsfunktion $G(h, P)$ von der Gestalt $|h - m(P)|^r$. $h(x)$ ist genau dann erwartungstreue Minimalschätzung, wenn für alle $P \in K$ das Fréchet-

Differential der Norm von $h(x) - m(P)$ für alle Zuwächse $v(x) \in N^r$ verschwindet [10].

Speziell gilt also: $E(v(x)h(x); P) = 0$ für alle $v(x) \in N^2$ und alle $P \in K$ charakterisiert die minimalen $h(x) \in H \subset L^2(P)$ mit Gewichtsfunktion $(h - m(P))^2$ [9].

Daraus folgt unmittelbar, daß für $r = 2$ mit je zwei Funktionalen $m_1(P)$, $m_2(P)$ über K, welche erwartungstreue Minimalschätzungen gestatten auch jede Linearkombination $\lambda_1 m_1(P) + \lambda_2 m_2(P)$ mit beliebigen reellen λ_1, λ_2 minimal geschätzt werden kann.

Für die tatsächliche Konstruktion erwartungstreuer Minimalschätzungen ist folgende Definition wichtig: Sei $T(x)$ eine Abbildung von R in eine Menge V. Zu jedem $P \in K$ erhalten wir durch die Abbildung $T(x)$ das induzierte Maß P^T. Dieses ist über einem meßbaren Raum $(V, \mathfrak{S})$ definiert, wobei $\mathfrak{S}$ mindestens jene Teilmengen $S \in V$ enthält, deren Urbild $T^{-1}(S)$ zu $\mathfrak{B}$ gehört. Es ist dann definitionsgemäß $P^T(S) = P(T^{-1}(S))$. Die Menge aller P^T bezeichnen wir mit K^T. Die bedingte Wahrscheinlichkeit $W(B \mid T; P)$ ist dann bekanntlich für jedes $B \in \mathfrak{B}$ nach dem Satz von Radon-Nikodym durch die Gültigkeit der Gleichung

$$W(B \cap T^{-1}(S);\ P) = \int_S W(B \mid T; P)\, dP^T \tag{5}$$

für alle $S \in \mathfrak{S}$ bis auf P^T-Nullmengen als Funktion über V definiert.

Man nennt nun $T(x)$ *erschöpfend* oder *erschöpfende Schätzung* für die Klasse K, wenn es für alle $B \in \mathfrak{B}$ eine über M definierte Funktion $W(B \mid T)$ gibt, welche nicht von $P \in K$ abhängt und für jedes $B \in \mathfrak{B}$ und alle $P \in K$ die Gleichung (5) für alle $S \in \mathfrak{S}$ (bis auf P^T-Nullmengen) erfüllt [11].

Es gilt nun in leichter Verallgemeinerung eines Satzes von Blackwell [12]:

Satz 3: Sei $G_1(z)$ konvex und nicht negativ und $G(h, P) = G_1(h - m(P))$. Es existiere für K eine erschöpfende Schätzung $T(x)$. Dann ist der bedingte Erwartungswert $E(h(x) \mid T)$, der nicht von $P \in K$ abhängt, erwartungstreu für $m(P)$ und nicht schlechter als $h(x) \in H$. Ist $G_1(z)$ strikt konvex, dann ist $E(h(x) \mid T)$ genau dann besser als $h(x)$, wenn $h(x)$ nicht von der Form $s(T(x))$ ist. Wenn $h(x)$ die Menge H durchläuft, bilden die $E(h(x) \mid T)$ eine vollständige Klasse Q.

Wir geben folgende wichtige Definition. Eine Klasse K von Wahrscheinlichkeitsmaßen heißt *vollständig* [9], wenn die (oben definierte) Menge N^1 nur aus $v(x) \equiv 0$ besteht.

Es gilt dann

Satz 4: Es seien die Voraussetzungen des Satzes 3 erfüllt. Überdies sei die Menge der durch $T(x)$ induzierten Wahrscheinlichkeitsmaße K^T vollständig. Dann existiert genau eine erwartungstreue Minimalschätzung [9].

Wenn die Voraussetzungen des Satzes 4 realisiert sind, genügt es für die Auffindung der erwartungstreuen Minimalschätzung, irgendeine Funktion $h(x) \in H$ zu kennen. Der bedingte Erwartungswert $E(h(x) \mid T)$ ist notwendig Minimalschätzung [12a].

Wir gehen nun von der Annahme aus, daß $(R, \mathfrak{B})$ von der Gestalt $(M^{(n)}, \mathfrak{S}^{(n)})$ und K von der Gestalt $K^{(n)}$ ist und kommen durch Anwendung der bisher erhaltenen Sätze zu einer Reihe von Ergebnissen, die vielfach als eigentlich „nichtparametrisch" bezeichnet werden. Die Elemente von $M^{(n)}$ sind von der Gestalt $(x_1, \ldots, x_n)$ mit

$$x_i \in M, \; i = 1, \ldots, n.$$

Ohne Verwechslungen befürchten zu müssen, bezeichnen wir diese Elemente von $M^{(n)}$ auch kurz mit x.

Es sei $t(x)$ jene Abbildung, welche jedem geordneten n-Tupel $(x_1, \ldots, x_n) \in M^{(n)}$ die Menge $\{x_1, \ldots, x_n\}$ zuordnet. Wir bezeichnen $t(x)$ als *Ordnungsschätzung*.

Wir bemerken, daß es sich bei dieser Abbildung um eine Verallgemeinerung des bekannten Begriffes der Ordnungsschätzungen handelt. Betrachten wir nämlich speziell den R_n und definieren wir über dem R_n n Funktionen $z_i(x_1, \ldots, x_n)$, $i = 1, \ldots, n$ in folgender Weise: Sei für $(x_1, \ldots, x_n)$ $x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_n}$, wobei $i_1, \ldots, i_n$ natürlich eine Permutation von $1, \ldots, n$ ist. Dann soll $z_j(x_1, \ldots, x_n) = x_{i_j}$ $j = 1, \ldots, n$ sein. Man nennt z_i die i-te Ordnungsschätzung und es ist leicht zu sehen, daß für $(x_1, \ldots, x_n) \in R_n$ die vorhin definierte Abbildung $t(x_1, \ldots, x_n)$ und $(x_1, \ldots, x_n) \to (x_{i_1}, \ldots, x_{i_n})$ dieselben Urbilder besitzen.

Wir kehren wieder zum allgemeinen Fall zurück und formulieren den

Satz 5: $t(x)$ ist für jede beliebige Klasse $K^{(n)}$ von Wahrscheinlichkeitsmaßen erschöpfend [13].

Wir bezeichnen jede beliebige über $M^{(n)}$ definierte Abbildung als *symmetrisch*, wenn sie gegenüber allen Permutationen ihrer Argumente invariant bleibt.

Es sei nun $h(x_1, \ldots, x_n)$ eine über $M^{(n)}$ definierte meßbare Funktion, für die $E(h(x_1, \ldots, x_n); P)$ für jedes $P \in K$ existiert. Dabei ist

$$E(h(x_1, \ldots, x_n); P)$$

eine abkürzende Schreibweise für

$$\int_M \dots \int_M h(x_1, \dots, x_n)\, dP \dots dP.$$

Weiter sei $\psi(t) = E(h(x_1, \dots, x_n) \mid t)$ und dieser bedingte Erwartungswert, der über der Menge aller Mengen der Form $\{x_1, \dots, x_n\}$ definiert ist, hängt nach Satz 5 von $P \in K$ nicht ab. Es ist leicht einzusehen, daß die über $M^{(n)}$ definierte Funktion $\psi(t(x))$ symmetrisch ist.

Diese Bemerkung liefert zusammen mit Satz 3 und Satz 5 den

Satz 6: Es sei $m(P)$ ein über K definiertes Funktional. Wenn mindestens eine erwartungstreue Schätzung $h(x_1, \dots, x_n)$ über $M^{(n)}$ für $m(P)$ existiert, das heißt,

$$\int_M \dots \int_M h(x_1, \dots, x_n)\, dP \dots dP = m(P) \quad \text{für alle} \quad P \in K,$$

dann ist die Klasse der erwartungstreuen und symmetrischen Schätzungen für $m(P)$ nicht leer und bildet unter der Voraussetzung einer nicht negativen und strikt konvexen Gewichtsfunktion eine vollständige Klasse Q.

Auf den naheliegenden Zusammenhang der Sätze 5 und 6 mit dem Invarianzprinzip sei hier nur kurz hingewiesen.

Wir bemerken nebenbei, daß jedes Funktional $m(P)$, welches in der Gestalt

$$\int_M \dots \int_M h(x_1, \dots, x_n)\, dP \dots dP$$

darstellbar ist, im Sinne von Mazur und Orlicz [14] eine *polynomische Operation* darstellt. Ist K konvex, dann ist dieses Funktional durch seine Werte für $n + 1$ Wahrscheinlichkeitsmaße $P \in K$, die nicht voneinander linear abhängig sind, über K eindeutig festgelegt.

Wir wollen nun noch spezieller annehmen, daß die Menge M der R_1 also die Produktmenge $M^{(n)}$ der R_n ist. $K^{(n)}$ sei insbesondere von der Gestalt $V_a^{(n)}$ oder $V_d^{(n)}$. Dann gilt

Satz 7: Legt man die Klasse $V_a^{(n)}$ oder $V_d^{(n)}$ von Wahrscheinlichkeitsmaßen zugrunde, dann ist die entsprechende Klasse der durch die oben definierte Abbildung $t(x)$ induzierten Maße vollständig [15].

Satz 7 besagt im wesentlichen, daß für gewisse Klassen von Wahrscheinlichkeitsmaßen die o nur trivial geschätzt werden kann. Legen wir aber die Klasse ${}_p\mathfrak{V}^{(n)}$ zugrunde, dann gibt es nicht triviale symmetrische Schätzungen für das über dieser Klasse identisch verschwindende Funktional. Man hat nämlich den [16]

Satz 8: Die über dem R_n definierte symmetrische Funktion

$$h(x_1, \dots, x_n)$$

ist genau dann erwartungstreue Schätzfunktion für das über ${}_p\mathfrak{F}^{(n)}$ definierte identisch verschwindende Funktional, wenn sie von der Gestalt $\sum_{i=1}^{n} \alpha(x_i)\, g(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ ist. Dabei ist $g(x_1, \ldots, x_{n-1})$ eine beliebige, über dem R_{n-1} definierte symmetrische beschränkte und meßbare Funktion und $\alpha(x)$ durch

$$\alpha(x) = \begin{cases} -(1-p) & x > 0 \\ p & x < 0 \end{cases} \quad \text{definiert.}$$

Unter Berücksichtigung von Satz 4, 5 und 7 erhalten wir den

Satz 9: Sei $K^{(n)}$ von der Gestalt $V_a^{(n)}$ oder $V_d^{(n)}$. Wenn mindestens eine erwartungstreue Schätzung für $m(P)$ existiert, dann gibt es genau eine notwendig symmetrische erwartungstreue Minimalschätzung für $m(P)$ mit nicht negativer und strikt konvexer Gewichtsfunktion.

Daraus folgt z. B., daß $\bar{x} = (x_1 + \cdots + x_n)/n$ die einzige erwartungstreue Minimalschätzung für das erste Moment aller absolut stetigen oder aller diskreten Verteilungen ist, soweit es existiert.

Wir bemerken, daß alle diese Sätze eine Konvexitätsforderung für die Gewichtsfunktion zur Voraussetzung haben. Diese Voraussetzung ist aber in gewissem Sinne unentbehrlich. Basu [17] hat nämlich gezeigt, daß nicht einmal im klassischen Fall der Schätzung des Mittelwertes einer Normalverteilung eine erwartungstreue Minimalschätzung existiert, wenn man gewisse nicht konvexe Gewichtsfunktionen zugrundelegt.

Wir machen noch einige kurze Angaben über die Verallgemeinerung dieser Definitionen und Sätze auf vektorwertige Funktionale $m(P)$, die Werte aus einem R_s ($s \geq 1$) annehmen. Es sei für $P \in K$

$$m(P) = (m_1(P), \ldots, m_s(P)).$$

Die über R definierte vektorwertige Funktion $h(x) = (h_1(x), \ldots, h_s(x))$ mit reellwertigen und meßbaren $h_i(x)$ heißt erwartungstreue Schätzfunktion für $m(P)$, wenn

$$E(h_i(x); P) = m_i(P) \quad \text{für} \quad i = 1, \ldots, s \quad \text{und alle} \quad P \in K$$

gilt. Wir schreiben dafür kürzer

$$E(h(x); P) = m(P). \tag{6}$$

Die Klasse aller durch (6) definierten $h(x)$ setzen wir als nicht leer voraus und bezeichnen sie wieder mit H.

Es seien $\lambda_1, \ldots, \lambda_s$ beliebige reelle Zahlen. Aus (6) folgt sofort, daß $\sum_{i=1}^{s} \lambda_i h_i(x)$ erwartungstreue Schätzfunktion für $\sum_{i=1}^{s} \lambda_i m_i(P)$ ist. Ist umgekehrt $\sum_{i=1}^{s} \lambda_i h_i(x)$ erwartungstreue Schätzfunktion für $\sum_{i=1}^{s} \lambda_i m_i(P)$ für alle s-Tupel reeller Zahlen $(\lambda_1, \ldots, \lambda_s)$, dann gilt (6).

Es liegt daher nahe, den Begriff der Minimalschätzung für vektorwertige Funktionale durch Betrachtung der reellwertigen Funktion $\sum_{i=1}^{s} \lambda_i h_i(x)$ auf die früheren Überlegungen zurückzuführen. Das geschieht bei Beschränkung auf quadratische Gewichtsfunktionen (so daß die Risikofunktion die Streuung darstellt) in der Weise, daß man mit der Bezeichnung

$$E[(h_i(x) - m_i(P))\,(h_j(x) - m_j(P)); P] = \sigma_{ij}^{h}(P), \quad i, j = 1, \ldots, s$$

die quadratische Form

$$\sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{ij}^{h}(P)$$

betrachtet. $h(x) \in H$ wird als Minimalschätzung bezeichnet, wenn für alle reellen $\lambda_1, \ldots, \lambda_s$ und für alle $\bar{h}(x) \in H$ und jedes $P \in K$

$$\sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{ij}^{h}(P) \leq \sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{ij}^{\bar{h}}(P) \tag{7}$$

ist. Setzt man die Matrix der $\sigma_{ij}^{h}(P)$ für alle $P \in K$ als positiv definit voraus, dann sind es auch die Matrizen der $\sigma_{ij}^{\bar{h}}(P)$ für alle $\bar{h}$. Bezeichnen wir mit $\sigma_{h}^{ij}(P)$ und $\sigma_{\bar{h}}^{ij}(P)$ die Elemente der inversen Matrizen, dann wird (meist algebraisch) bewiesen, daß aus (7)

$$\sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{h}^{ij}(P) \geq \sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{\bar{h}}^{ij}(P) \tag{8}$$

für alle reellen λ_i, $\bar{h} \in H$ und $P \in K$ folgt und daraus geschlossen, daß für jede Minimalschätzung $h(x) \in H$ das Konzentrationsellipsoid

$$\sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{h}^{ij}(P) = 2s + 2$$

für jedes $h \in H$ und alle $P \in K$ stets innerhalb des Konzentrationsellipsoides

$$\sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{\bar{h}}^{ij}(P) = 2s + 2$$

liegt [18].

Wir bemerken übrigens, daß (8) aus (7) ohne Rechnung folgt, da für jedes $P \in K$ die durch $\sum_{j,i=1}^{s} \lambda_i \lambda_j \sigma_h^{ij}(P) \leq 1$, bzw. $\sum_{i,j=1}^{s} \lambda_i \lambda_j \sigma_{ij}^{h}(P) \leq 1$ definierten konvexen Gebilde zueinander polar sind. Man kann daher ganz analoge Überlegungen für allgemeinere konvexe Gewichtsfunktionen anstellen.

Diese Untersuchungen sind besonders dann von Nutzen, wenn die Existenz einer für H erschöpfenden Schätzfunktion $T(x)$ bekannt ist. Es gilt dann nach Satz 3 mit der Bezeichnung $E(h_i(x) \mid T) = \psi_i(T)$ für jede konvexe Gewichtsfunktion

$$E\left(G\left(\sum_{i=1}^{s} \lambda_i (h_i - m_i(P))\right); P\right) \geq E\left(G\left(\sum_{i=1}^{s} \lambda_i (\psi_i - m_i(P))\right); P\right)$$

für alle $P \in K$. Daraus lassen sich dann wie früher eine Reihe wichtiger Folgerungen ziehen.

Mit der Theorie der erwartungstreuen Schätzungen hängt die Theorie der *Toleranzbereiche* eng zusammen. Wir gehen von einem meßbaren Raum $(R, \mathfrak{B})$ aus. Es sei $Q(x)$ eine Abbildung von R in $\mathfrak{B}$. Für jedes $P \in K$ ist $P(Q(x))$ eine reellwertige Funktion, welche wir als meßbar voraussetzen, d. h. eine zufällige Variable. Wir bezeichnen das durch $P(Q(x))$ auf R_1 induzierte Wahrscheinlichkeitsmaß mit W_P. $Q(x)$ heiße nichtparametrischer Toleranzbereich für K, wenn das induzierte Wahrscheinlichkeitsmaß W_P für alle $P \in K$ dasselbe ist.

Wenn R ein Produktraum von der Gestalt $M^{(n)}$ und K eine dementsprechende Klasse von Produktmaßen $K^{(n)}$ ist, dann betrachten wir Abbildungen $Q(x_1, \ldots, x_n)$, $x_i \in M$ von $M^{(n)}$ in $\mathfrak{S}$. Die Definition des nichtparametrischen Toleranzbereiches gründet sich dann auf der Betrachtung von $P(Q(x_1, \ldots, x_n))$ mit $P \in K$.

Es sei nun $c_{x_1, \ldots, x_n}(y)$ die charakteristische Funktion von $Q(x_1, \ldots, x_n)$. Dann gilt der

Satz 10: [19] $Q(x_1, \ldots, x_n)$ ist genau dann nichtparametrischer Toleranzbereich, wenn eine Folge nicht negativer reeller Zahlen m_i mit $m_i \leq 1$ existiert, so daß für jedes $s \geq 1$

$$\prod_{j=1}^{s} c_{x_1, \ldots, x_n}(y_j) - m_s$$

erwartungstreue Schätzung für das identisch verschwindende Funktional über $K^{(n+s)}$ ist. Überdies gilt

$$m_i = \int_0^1 p^i \, \mathrm{d}\, W_P(p) \quad \text{für alle} \quad P \in K.$$

Die Konstruktion nichtparametrischer Toleranzbereiche ist im R_n unter sehr allgemeinen Voraussetzungen möglich und basiert auf der Betrachtung der geordneten Schätzfunktionen. Wir können jedoch hierauf nicht eingehen [20].

III. Wir betrachten jetzt neben der Klasse K, von der wir voraussetzen, daß sie mindestens zwei Wahrscheinlichkeitsverteilungen enthält, eine nicht leere Teilklasse k, welche wir die Nullhypothese nennen. Wir passen den Raum C der Entscheidungen unserem Problem an und nehmen an, daß er nur aus zwei Elementen c_1, c_2 besteht, die man als „Annehmen bzw. Ablehnen der Nullhypothese" interpretieren kann. Ebenso spezialisieren wir die Gewichtsfunktion. Es sei

$$d_i > 0, \quad i = 1, 2$$

und

$$(9) \qquad G(c_1, P) = \begin{cases} 0 \\ d_1 \end{cases} \qquad G(c_2, P) = \begin{cases} d_2 & \text{für} \quad P \in k \\ 0 & \text{für} \quad P \in K \end{cases}$$

Die Entscheidungsfunktionen $\varepsilon(x)$ sind unter diesen speziellen Voraussetzungen eindeutig festgelegt, wenn etwa die Wahrscheinlichkeit $\varphi(x)$ für die Ablehnung der Nullhypothese vorgegeben ist. Wir geben in diesem Sinne folgende Definition: Jede meßbare Abbildung $\varphi(x)$ von R in $[0,1]$ ist ein *Test*.

Wir erhalten aus (9) für die Risikofunktion

$$(10) \qquad R(\varphi, P) = \begin{cases} d_2 E(\varphi; P) & \text{für } P \in k \\ d_1 (1 - E(\varphi; P)) & \text{für } P \in K - k \end{cases}$$

In der klassischen Testtheorie, die in ihrer ursprünglichen Form auf Neyman und Pearson [21] zurückgeht, gibt man eine reelle Zahl α mit $0 < \alpha < 1$ vor, die sogenannte Sicherheitswahrscheinlichkeit, und betrachtet nur solche Tests $\varphi(x)$, für die

$$(11) \qquad \sup_{P \in k} E(\varphi; P) = \alpha$$

Man betrachtet also nur solche Tests, welche die richtige Nullhypothese „im Mittel" höchstens mit Wahrscheinlichkeit α ablehnen. Die Klasse der Tests, welche der Bedingung (11) genügen, bezeichnen wir mit Φ_α. Für festes φ möge $E(\varphi; P)$ für $P \in K - k$ als *Gütefunktional* des Tests φ bezeichnet werden.

Gilt insbesondere $E(\varphi; P) \equiv \alpha$ für $P \in k$, dann bezeichnet man $\varphi(x)$ als *ähnlich* zum Stichprobenraum R.

Zunächst interessieren uns wieder Minimaltests. Wir stellen uns also die Aufgabe, die Risikofunktion (10) für die $\varphi \in \Phi_\alpha$ bezüglich $K - k$

zu minimisieren. Wegen (10) ist dies damit gleichbedeutend, daß das Gütefunktional unter der Bedingung (11) maximisiert werden soll. Tests, die das leisten, sollen *trennscharf* bezüglich $K - k$ genannt werden. Trennscharfe Tests existieren nur in Ausnahmefällen. Es gilt jedoch der

Satz 11: K bestehe aus genau zwei Wahrscheinlichkeitsmaßen. Dann existiert stets ein trennscharfer Test [22].

Nach Wald und Dantzig [23] sind solche trennscharfe Tests „im allgemeinen" eindeutig bestimmt, eine Folge der Abgeschlossenheit und Konvexität des Wertevorrates einer beschränkten vektorwertigen Maßfunktion mit Werten aus einem R_n.

Wald [24] hat den Begriff des trennscharfen Tests zu dem des *strengen Tests* verallgemeinert. Es sei

$$S(P) = \sup_{\varphi \in \Phi_\alpha} E(\varphi; P) \quad \text{für} \quad P \in K - k$$

$\varphi_0 \in \Phi_\alpha$ heißt streng bezüglich $K - k$, wenn

$$\sup_{P \in K-k} S(P) - E(\varphi_0; P) \leq \sup_{P \in K-k} S(P) - E(\varphi; P) \quad \text{für alle} \quad \varphi \in \Phi_\alpha .$$

Aus dieser Definition folgt sofort: Wenn $S(P)$ auf $K - k$ konstant ist, dann ist φ_0 minimax bezüglich aller $\varphi \in \Phi_\alpha$.

Bedeutsam ist der Zusammenhang des Konzepts der strengen Tests mit dem *Invarianzprinzip*, das für Testprobleme mit der Gewichtsfunktion (9) folgende speziellere Gestalt besitzen soll: γ_K sei ein Homomorphismus von $\mathfrak{G}$ in die Menge aller eineindeutigen Transformationen von K in sich, welche k und $K - k$ in sich überführen. γ_C bildet $\mathfrak{G}$ auf die Identität ab. Die Invarianz der Entscheidungsfunktion drückt sich dann durch

$$\varphi(g_K x) = \varphi(x) \tag{12}$$

aus. Die Gruppe $\mathfrak{G}$ nennen wir für das Testproblem zulässig. Insbesondere wollen wir diese Definitionen unter der Voraussetzung benützen, daß ein für alle $P \in K$ dominantes Maß μ existiert, so daß nach dem Satz von Radon-Nikodym alle P bezüglich μ Dichten besitzen. Wir nennen in diesem Fall eine Entscheidungsfunktion $\varphi(x)$ auch invariant, wenn (12) für alle x bis auf μ-Nullmengen gilt. Hängt diese Nullmenge vom Element g der Gruppe $\mathfrak{G}$ ab, dann nennen wir den Test $\varphi(x)$ fast invariant. Falls die Klasse der fast invarianten (invarianten) Tests aus Φ_α nicht leer ist, bezeichnen wir sie mit I'_α (I_α). Es gilt der

Satz 12: [25]: Es existiere ein für alle $P \in K$ dominantes Maß μ. Es sei $\mathfrak{G}$ eine für ein Testproblem zulässige Gruppe. Überdies sei ein

σ-Körper $\mathfrak{A}$ von Teilmengen aus $\mathfrak{G}$ definiert und über $\mathfrak{A}$ eine Folge asymptotisch invarianter Wahrscheinlichkeitsmaße q_n gegeben, d. h. es soll

$$\lim_{n\to\infty}(q_n(gA) - q_n(A)) = 0 \quad \text{für alle} \quad g \in \mathfrak{G} \quad \text{und alle} \quad A \in \mathfrak{A}$$

sein [26]. Dann existiert ein φ_0 aus I'_α, welches $\inf_{P\in K-k} E(\varphi; P)$ für $\varphi \in \Phi_\alpha$ maximisiert, also ein invarianter Minimatest aus Φ_α.

Bei entsprechenden Voraussetzungen über Abzählbarkeitseigenschaften von $\mathfrak{G}$ kann man natürlich in der Aussage des Satzes (12) I'_α durch I_α ersetzen [27].

Der Beweis dieses wichtigen Satzes beruht im wesentlichen auf der schwachen Kompaktheit von Φ_α.

Aus Satz 12 folgert man leicht, daß unter entsprechenden Voraussetzungen jeder Test, der für die Klasse I_α bezüglich $K - k$ trennscharf ist, für die Gesamtheit aller Tests aus Φ_α streng ist. Die Bedeutung dieser Bemerkung liegt darin, daß $\mathfrak{G}$ $K - k$ in invariante Teilbereiche zerschlägt und daher unter Umständen die Auffindung invarianter trennscharfer Tests erleichtert wird, da die Wahrscheinlichkeitsverteilung jedes Tests aus I_α für alle P aus einem beliebigen invarianten Teilbereich von $K - k$ invariant bleibt. Wir betrachten nun einige spezielle Gruppen.

A. Wir legen den R_n zugrunde. Es sei $\mathfrak{G}^*$ die Gruppe aller stetigen und eigentlich monoton wachsenden Transformationen g^* des R_1 in sich. Wir definieren Transformationen g des R_n in sich, gemäß

$$gx = (g^*x_1, \ldots, g^*x_n).$$

Man sieht sofort, daß die g die Elemente einer Gruppe $\mathfrak{G}_A$ sind, welche auf die Punkte des R_n wirken, wenn g^* die Gruppe $\mathfrak{G}^*$ durchläuft.

B. Wir betrachten jene Untergruppe $\mathfrak{G}^*_{1/_2}$ von $\mathfrak{G}^*$, deren Elemente durch $g^*x > 0$ für $x > 0$, $g^*x < 0$ für $x < 0$ gekennzeichnet sind. Es seien $g_1^*, \ldots, g_n^*$ irgendwelche Elemente von $\mathfrak{G}^*_{1/_2}$. Wir definieren Transformationen g des R_n in sich gemäß

$$gx = (g_1^*x_1, \ldots, g_n^*x_n),$$

wobei $g_1^*, \ldots, g_n^*$ alle Elemente von $G^*_{1/_2}$ durchlaufen sollen. Wir bezeichnen die Gruppe der g mit $\mathfrak{G}_B$.

C. Wir betrachten nun die Elemente g_s^* jener Untergruppe $\mathfrak{G}_s^*$ von $\mathfrak{G}^*$, deren Elemente durch

$$g_s^*x = -g_s^* - x$$

gekennzeichnet sind. Überdies bezeichnen wir mit s die Spiegelung der

reellen Geraden am Nullpunkt: $sx = -x$. Die durch $\mathfrak{G}_s^*$ und s erzeugte Gruppe bezeichnen wir mit $\overline{\mathfrak{G}}_s^*$, ihre Elemente mit $\bar{g}_s^*$. Wir definieren Transformationen g des R_n in sich gemäß

$$gx = (\bar{g}_s^* x_1, \ldots, \bar{g}_s^* x_n).$$

Die Gesamtheit der Elemente g, welche wir mit $\mathfrak{G}_C$ bezeichnen, bilden eine Gruppe, wenn $\bar{g}_s^*$ die Gruppe $\overline{\mathfrak{G}}_s^*$ durchläuft.

A. Wir legen nun die Gruppe $\mathfrak{G}_A$ zugrunde. Es sei X die Menge der $x \in R_n$ mit $x_i \neq x_j$ für $i \neq j$, $i, j = 1, \ldots, n$. Für alle Wahrscheinlichkeitsmaße aus $S^{(n)}_{1,\ldots,n}$ ist die Menge $R_n - X$ eine Nullmenge. Für die $x \in X$ kann man die x_i eindeutig der Größe nach anordnen. Ist in dieser Anordnung x_i das r_i-te Element, dann sagt man x_i habe den Rang r_i. Für alle $x \in X$ ist die Abbildung $r(x)$, welche durch $x \to (r_1, \ldots, r_n)$ erklärt ist, wohl definiert. Die Abbildung $r(x)$ ist überdies invariant gegenüber $\mathfrak{G}_A$, d. h. es gilt für jede der $n!$ Mengen $N(r_1, \ldots, r_n) = \{x : r(x) = (r_1, \ldots, r_n)\}$, $\mathfrak{G}N \subseteq N$. Man überlegt sich leicht, daß jede gegenüber $\mathfrak{G}_A$ invariante über dem R_n definierte Abbildung von der Gestalt $s(r(x))$ ist.

Jede meßbare Abbildung von $r(x)$ in $[0,1]$ bezeichnen wir als *ranginvarianten Test.*

Für jedes $g \in \mathfrak{G}_A$ und $F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i) \in S^{(n)}_{1,\ldots,n}$ definieren wir

$$g_K F = \prod_{i=1}^{n} F_i(g^{*-1} x_i). \tag{13}$$

Dadurch wird ein Homomorphismus von $\mathfrak{G}_A$ in die Menge der eineindeutigen Transformationen von $S^{(n)}_{1,\ldots,n}$ in sich erklärt. Das Invarianzprinzip für Tests legt es nahe, Teilmengen von $S^{(n)}_{1,\ldots,n}$ aufzusuchen, welche durch das homomorphe Bild von $\mathfrak{G}_A$ in sich übergeführt werden [28]. Es gilt der

Satz 13: [29] Für $i = 1, \ldots, n$ seien stetige und monotone Funktionen $f_i(x)$ definiert, welche den Bedingungen

$$f_i(0) = 0, \; f_i(1) = 1 \tag{14}$$

genügen. Dann ist die Menge aller Wahrscheinlichkeitsmaße, welche durch Verteilungsfunktionen der Gestalt $\prod_{i=1}^{n} f_i(F(x_i))$ gegeben sind, invariant gegenüber $\mathfrak{G}_A$, wenn $F(x)$ S durchläuft.

Insbesondere ist also $E(\varphi(x); \prod_{i=1}^{n} f_i(F(x_i)))$ für alle $F(x) \in S$ und jeden ranginvarianten Test $\varphi(x)$ konstant. Sind also $f_i^{(j)}(x)$, $j = 1, 2$ und

$i = 1, \ldots, n$ Funktionen, welche von der im Satz 13 angegebenen Art sind und insbesondere (14) genügen und besteht k genau aus der Menge aller Wahrscheinlichkeitsmaße, welche durch $\prod_{i=1}^{n} f_i^{(1)}(F(x_i))$ definiert sind mit $F(x) \in S$ und ist $K - k$ entsprechend durch $\prod_{i=1}^{n} f_i^{(2)}(F(x_i))$ definiert, dann läßt sich nach Satz 11 stets ein für die Nullhypothese k ähnlicher und bezüglich $K - k$ trennscharfer Test konstruieren, welcher die Gruppe $\mathfrak{G}_A$ zuläßt.

Die Betrachtung des wichtigen Sonderfalles $f_i^{(1)}(x) \equiv x$ führt uns zu einem der wichtigsten Anwendungsgebiete der ranginvarianten Tests, zum „*Problem der zwei Stichproben*". Es sei $m, n \geq 1$. Wir nehmen an, daß die Nullhypothese k mit der Klasse $S^{(n+m)}$ übereinstimmt. Die Gegenhypothese $K - k$ enthalte alle Wahrscheinlichkeitsmaße, deren Verteilungsfunktionen von der Gestalt $\prod_{i=1}^{n} F(x_i) \prod_{j=1}^{m} G(x_{n+j})$ sind mit $F(x), G(x) \in S$, $F(x) \neq G(x)$. Ein Element des R_{n+m} bezeichnen wir jetzt zweckmäßigerweise durch $(x_1, \ldots, x_n\, y_1, \ldots, y_m)$ oder kurz mit (x, y). $(x_1, \ldots, x_n)$ soll nämlich das Ergebnis der ersten und $(y_1, \ldots, y_m)$ das Ergebnis der zweiten Stichprobe darstellen.

Um an Hand der beiden Stichproben zu einer Entscheidung über die Gültigkeit der Nullhypothese k zu gelangen, sind eine Reihe von ranginvarianten Zweistichprobentests konstruiert worden, so z. B. von Terry [30], Van der Waerden [31] und Wilcoxon [32]. Nach Satz 13 sind diese Tests für $k = S^{(n+m)}$ ähnlich zum Stichprobenraum. Wir bezeichnen bei der Abbildung $r(x, y)$ die Rangzahlen von $y_1, \ldots, y_m$ mit $s_1, \ldots, s_m$. Es sei $N = n + m$ und Z_{N, s_i} die s_i-te Ordnungsschätzung in einer Stichprobe normal verteilter zufälliger Variabler mit Mittelwert 0 und Streuung 1 vom Umfang N, das sind also N unabhängige zufällige Variable mit der angegebenen Verteilung. Dann wird der Test $\varphi(x, y)$ bei Terry unter Benützung von $\sum_{i=1}^{m} E(Z_{N, s_i})$ definiert. Es ist also bei vorgegebener Sicherheitswahrscheinlichkeit α und passend gewähltem $\varkappa_\alpha$ und c mit $0 < c < 1$

$$\varphi(x, y) = \begin{cases} 1 & \sum_{i=1}^{m} E(Z_{N, s_i}) < \varkappa_\alpha \\ c & \sum_{i=1}^{m} E(Z_{N, s_i}) = \varkappa_\alpha \\ 0 & \sum_{i=1}^{m} E(Z_{N, s_i}) > \varkappa_\alpha . \end{cases}$$

Bei Van der Waerden wird die Testfunktion in analoger Weise mittels $\sum_{i=1}^{m} \psi\left(\frac{s_i}{N+1}\right)$ konstruiert, wobei $\psi(q)$ das q-Quantil der standardisierten Normalverteilung bezeichnet.

Wilcoxon verwendet schließlich $\sum_{i=1}^{m} s_i$.

Für den Vergleich dieser und ähnlicher Tests für das Zweistichprobenproblem ist die Untersuchung des Gütefunktionals natürlich von besonderer Bedeutung. Erste Auskunft gibt natürlich Satz 13, doch ist damit für die oben angegebene allgemeine Alternative $K - k$ des Zweistichprobenproblems wenig geleistet. Von den neueren Untersuchungen der Gütefunktion für eine Reihe speziellerer Alternativen erwähnen wir eine Arbeit von Savage [33].

Interessante Probleme knüpfen sich an den Zweistichprobentest von Smirnow [34]. Hierzu definieren wir die empirische Verteilungsfunktion $F_n(z)$ einer endlichen Menge reeller Zahlen $z_1, z_2, \ldots, z_n$, welche der Bedingung $z_1 < z_2 < \cdots < z_n$ genügen. Es sei

$$F_n(z) = \begin{cases} 0 & z < z_1 \\ k/n & z_k \leq z < z_{k+1}, \quad k = 1, \ldots, n-1 \\ 1 & z \geq z_n \end{cases}$$

Wenn nun X die oben erläuterte Bedeutung hat und Y entsprechend für $(y_1, \ldots, y_m)$ definiert ist, dann hat $X \times Y$ für alle Wahrscheinlichkeitsmaße aus $S^{(n+m)}_{1,\ldots,(n+m)}$ das Maß 1 und jedem n-Tupel $(x_1, \ldots, x_n) \in X$ bzw. m-Tupel $(y_1, \ldots, y_m) \in Y$ entspricht eine empirische Verteilungsfunktion $F_n(x)$ bzw. $G_m(y)$. Der Zweistichprobentest von Smirnow beruht nun auf der Betrachtung von

$$\sup_{-\infty < z < \infty} |F_n(z) - G_m(z)|.$$

Es gilt der

Satz 14: [34] Sei $N = mn/(m+n)$. Dann ist für alle Wahrscheinlichkeitsmaße aus $S^{(n+m)}$

$$\lim_{\substack{m,n \to \infty \\ m/n \to a}} W\left(\sqrt{N} \sup_{-n < z < \infty} |F_n(z) - G_m(z)| \leq x\right) = \begin{cases} \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2x^2} & x > 0 \\ 0 & x < 0 \end{cases}$$

mit $a \neq 0$.

Dieser Satz sagt allerdings nur etwas über die asymptotische Verteilung der zufälligen Größe $\sup_{-\infty < z < \infty} |F_n(z) - G_m(z)|$ aus. Für unsere Zwecke ist besonders wichtig die Verteilung für endliche n, m zu kennen. Ergebnisse in dieser Richtung stammen z. B. von Koroljuk [35].

Zusammenfassende Angaben über diese und andere bisher erzielten Ergebnisse finden sich bei Gnedenko [36]. Überdies weisen wir auf einen Bericht von Darling hin [37].

B. Wir interessieren uns jetzt für Tests, welche gegenüber der Gruppe $\mathfrak{G}_B$ invariant bleiben. Hierzu definieren wir zunächst die Menge N der $x = (x_1, \ldots, x_n) \in R_n$, für die für mindestens ein $i (1 \leq i \leq n)$

$$x_i = 0$$

gilt. Offenbar ist für alle $P \in S^{(n)}_{1,\ldots,n}$ $P(N) = 0$. Wir definieren für alle $x \in R_n - N$ eine Abbildung $\varepsilon(x)$ durch $(x_1, \ldots, x_n) \to (\varepsilon_1, \ldots, \varepsilon_n)$, wobei $\varepsilon_i = \operatorname{sign} x_i = \pm 1$. Man sieht sofort, daß diese Abbildung gegenüber der Gruppe $\mathfrak{G}_B$ invariant ist und jede bezüglich $\mathfrak{G}_B$ invariante Abbildung von der Gestalt $s(\varepsilon(x))$ ist.

Wir bezeichnen jede meßbare Abbildung von $\varepsilon(x)$ in [0,1] als *Vorzeichentest*. Erklärt man einen Homomorphismus von $\mathfrak{G}_B$ in die Menge der Transformationen von $S^{(n)}_{1,\ldots,n}$ in sich nach dem Vorbild der Definition (13), dann sieht man leicht, daß die Menge ${}_{1/2}S^{(n)}_{1,\ldots,n}$ invariant bleibt. Somit ist jeder Vorzeichentest ähnlich für die Nullhypothese $k = {}_{1/2}S^{(n)}_{1,\ldots,n}$. In diesem Sinne kann man die Vorzeichentests als *Mediantests* bezeichnen.

Der einfachste Vorzeichentest ist der sogenannte *Zeichentest*, welcher den Test mittels der Größe $\varepsilon_1 + \cdots + \varepsilon_n$ konstruiert.

Die Vorzeichentests und insbesondere der Zeichentest werden gelegentlich auch als „*Symmetrietests*" verwendet, d. h. man geht von der Nullhypothese $k = {}_sS^{(n)}_{1,\ldots,n}$ aus. Wir haben aber eben erwähnt, daß die Vorzeichentests für eine umfassendere Nullhypothese ähnlich sind.

C. Symmetrietests stehen in engem Zusammenhang mit der Gruppe $\mathfrak{G}_C$. Wir betrachten die Menge $\overline{X} \in R_n$, deren Elemente $(x_1, \ldots, x_n)$ durch $|x_i| \neq |x_j|$ für $i \neq j$; $i, j = 1, \ldots, n$ definiert sind. Für die $x \in \overline{X}$ ist die Abbildung $b(x) : x \to (s_1, \ldots, s_n)$ wohldefiniert, wobei s_i die Rangzahl von $|x_i|$ ist. Die Abbildung $b(x)$ ist gegenüber der Gruppe $\mathfrak{G}_C$ invariant und jede invariante Abbildung ist von der Form $s(b(x))$. Jede meßbare Abbildung von $b(x)$ in [0,1] bezeichnen wir als Symmetrietest. Man sieht nämlich wieder leicht, daß, kurz gesagt, die Gruppe $\mathfrak{G}_C$ die Menge ${}_sS^{(n)}_{1,\ldots,n}$ in sich überführt. Es sind also alle Symmetrietests ähnlich für die Nullhypothese $k = {}_sS^{(n)}_{1,\ldots,n}$.

In neuerer Zeit sind viele Symmetrietests konstruiert worden, und zwar meist für die umfassendere Nullhypothese $k = {}_s\mathfrak{F}^{(n)}_{1,\ldots,n}$. Wir weisen auf die Untersuchungen von Hemelrijk [38] und Ruist [39] hin. Genauer wollen wir auf einen kürzlich von van Eeden und Benard [40] beschriebenen Symmetrietest eingehen, welcher als Spe-

zialfälle z. B. den Symmetrietest von Fisher und den von Wilcoxon enthält. Es werde für jedes $(x_1, \ldots, x_n) \in R_n$ definiert: Die Anzahl a_i der x_j, welche denselben Wert u_i besitzen und die Anzahl b_i der x_j, welche denselben Wert $-u_i$ haben, wobei

$$0 < u_1 < \cdots < u_k, \quad i = 1, \ldots, k$$

und $0 \leq k \leq n$ gelten. ($k = 0$ soll genau dann gelten, wenn alle x_j verschwinden.) Es sei für $1 \leq i \leq k$, $c_i = a_i + b_i$ und $r_i = \sum_{j=1}^{i} c_j - {}^1/_2(c_i - 1)$ r_i kann man als Rangzahl jener Größen x_j auffassen, deren Betrag gleich u_i ist. Sei I die Menge $\{1, \ldots, n\}$ und $\omega(u, r)$ eine über $R_1 \times I$ definierte und für jedes r in u meßbare Funktion. Sei $\omega_i = \omega(u_i, r_i)$. Für gegebenes $k, c_1, \ldots, c_k, u_1, \ldots, u_k$ wird die Größe

$$T = \sum_{i=1}^{k} \omega_i \cdot (a_i - b_i)$$

betrachtet. Auf ihr basiert die Konstruktion des Symmetrietests. Man zeigt nämlich, daß die bedingte Verteilung von T unter der Hypothese $\{k, c_i, u_i, (1 \leq i \leq k)\}$ unabhängig davon ist, welches Wahrscheinlichkeitsmaß aus ${}_s\mathfrak{F}^{(n)}_{1,\ldots,n}$ zugrunde liegt. Überdies gehört diese Verteilung zu ${}_s\mathfrak{F}$.

Für $\omega(u, r) = u$ erhält man den Test von Fisher, für $\omega(u, r) = r$ den von Wilcoxon.

Wir wenden uns nun noch einer ganz anderen Frage zu. Wir haben in I. den Begriff der Unverfälschtheit einer Entscheidungsfunktion erklärt. Spezialisiert man diesen Begriff auf Tests, dann ergibt sich mit $\alpha = \frac{d_1}{d_1 + d_2}$, wobei d_1, d_2 gemäß (9) erklärt sind

$$\begin{aligned} &E(\varphi(x); P) \leq \alpha \quad \text{für} \quad P \in k \\ &E(\varphi(x); P) \geq \alpha \quad \text{für} \quad P \in K - k. \end{aligned} \tag{15}$$

Die Bedingung (15) ist besonders dann von Interesse, wenn für mindestens ein $P \in K - k$ das Größerzeichen gilt. Die Forderung der Unverfälschtheit für einen Test ist besonders in praktischer Hinsicht von großer Bedeutung. Ist nämlich die zweite Ungleichung (15) für einen Test $\bar{\varphi}(x)$ für kein $P \in K - k$ erfüllt, dann wähle man an dessen Stelle $\varphi(x) \equiv \alpha$. (15) ist dann trivialerweise erfüllt. Der Test ist aber dann von jedem Experiment unabhängig und trotzdem „günstiger" als $\bar{\varphi}(x)$.

Von Walter [41] stammt nun die interessante Bemerkung, daß es ranginvariante Tests für das Zweistichprobenproblem gibt, für die in

der zweiten Ungleichung (15) überall das Größerzeichen gilt. Ersetzt man aber $\varphi(x)$ durch $1-\varphi(x)$, dann erhält man offenbar ranginvariante Tests, für die in der zweiten Ungleichung (15) überall das Kleinerzeichen steht.

Zwischen den unverfälschten Tests und den Minimaxtests besteht ein Zusammenhang, den wir in der Formulierung von Sverdrup wiedergeben [42]:

Satz 15: Für alle $P \in K$ existiere ein dominantes Maß μ, so, daß jedem P eine Dichte bzg. μ zugeordnet werden kann. Eine Folge von Maßen aus K heiße konvergent gegen ein Maß aus K, wenn die zugehörigen Dichten für alle $x \in R$ punktweise gegen eine Dichte konvergieren. Dadurch wird eine Topologie in K induziert. Es möge eine nichtleere Teilmenge aus K geben, die aus Häufungspunkten von $K - k$ (k die Nullhypothese) besteht. Dann ist jeder unverfälschte Test für die Nullhypothese k und Alternative $K - k$ mit $\alpha = \frac{d_1}{d_1 + d_2}$ Minimaxtest.

Wir schließen mit folgender Bemerkung: In gewissem Sinne sind die nichtparametrischen Tests dadurch gekennzeichnet, daß einer „umfangreichen" Nullhypothese eine „umfangreiche" Alternative gegenübersteht. Allgemeiner liegen den nichtparametrischen Methoden „umfangreiche" Klassen K von Wahrscheinlichkeitsmaßen zugrunde. Die Frage liegt nahe, ob z. B. ein nichtparametrischer Test stets „im Stande ist" zwischen so „umfangreichen" Hypothesen zu unterscheiden. Untersuchungen in dieser Richtung wurden kürzlich mit negativem Ergebnis durchgeführt. Es gilt nämlich der

Satz 16: [43] Es sei K eine konvexe Klasse eindimensionaler Wahrscheinlichkeitsmaße, so daß für alle $P \in K$ $\int_{R_1} x\,dP$ stets existiert. Überdies existiere zu jedem reellen m ein $P \in K$, so daß

$$\int_{R_1} x\,dP = m \tag{16}$$

gilt. Es sei K_m die Teilmenge von K, für die (16) für ein festes m gilt. Für jede beschränkte und meßbare Funktion $\varphi(x)$, welche über dem R_n definiert ist, sind $\inf_{P^{(n)} \in K_m^{(n)}} E(\varphi(x); P)$ und $\sup_{P^{(n)} \in K_m^{(n)}} E(\varphi(x); P)$ unabhängig von m.

Hier spiegelt sich der einfache Sachverhalt wider, daß „kleine" Änderungen von Verteilungen „große" Schwankungen des Erwartnugswertes zur Folge haben können.

Literaturverzeichnis

[o] Zusatz bei der Korrektur (25. 2. 1959): Asymptotische Probleme erfreuen sich zunehmender Bedeutung. Vgl. etwa H. Chernoff und I. R. Savage: Ann. math. Statistics **29**, 972—994 (1958) und die dort angegebene Literatur.

[1] Für alle der Wahrscheinlichkeitstheorie entnommenen Begriffe vgl. man z. B. A. Kolmogorow, Grundbegriffe der Wahrscheinlichkeitsrechnung (Ergebn. der Math. u. ihrer Grenzgebiete). Berlin 1933, Springer-Verlag oder H. Richter, Wahrscheinlichkeitstheorie. Berlin-Göttingen-Heidelberg 1956, Springer-Verlag.

[2] Vgl. E. Ruist. Ark. Mat. **3**, 133—163 (1955). Eine eingehende Analyse des Begriffes „nichtparametrisch" findet sich in der Arbeit von M. G. Kendall und R. M. Sundrum. Revue Inst. internat. Statist. **21**, 124—134 (1953).

[3] $\mathfrak{S}^{(n)}$ soll natürlich den durch die Mengen der Gestalt $\prod_{i=1}^{n} A_i$ mit $A_i \varepsilon \mathfrak{S}$ erzeugten σ-Körper bedeuten.

[4] Dieser Begriff und die damit im Zusammenhang stehenden Definitionen gehen auf A. Wald zurück.

[5] E. L. Lehmann. Ann. math. Statistics **22**, 587—592 (1951).

[6] Vgl. hierzu J. Kiefer. Ann. math. Statistics **28**, 573—601 (1957). und die dort angegebene Literatur.

[7] F. N. David und J. Neyman. Statistical Research Mem. **2**, 105—116 (1938).

[8] A. Chinčin. C. r. Acad. Sci., Paris **188**, 477—479 (1929).

[9] E. L. Lehmann und H. Scheffé. Sankhya **10**, 305—339 (1950).

[10] L. Schmetterer. Mitteil.-Bl. math. Statistik **9**, 147—152 (1957).

[11] Zur Definition der erschöpfenden Schätzungen und wichtigen Eigenschaften derselben vgl. man R. P. Halmos und L. J. Savage. Ann. math. Statistics **20**, 225—241 (1949).

[12] D. Blackwell. Ann. math. Statistics **18**, 105—110 (1947). Vgl. auch A. N. Kolmogorow. Izvestija Akad. Nauk SSSR, Ser. mat. **14**, 303-326 (1950).

[12a] Zusatz bei der Korrektur (25. 2. 1959): Unter gewissen Bedingungen ist neuerdings von R. R. Bahadur, Sankhya **18**, 211—223 (1957) für den Fall $G_1(z) = z^2$ eine Umkehrung des wichtigen Satzes 4 gefunden worden.

[13] Vgl. z. B. D. A. S. Frazer, Nonparametric Methods in Statistics. New York-London 1957, John Wiley & Sons-Chapman & Hall, 141.

[14] S. Mazur und W. Orlicz. Studia math. **5**, 50—68 (1934).

[15] R. P. Halmos. Ann. math. Statistics **17**, 34—43 (1946) und D. A. S. Frazer. Canadian J. Math. **6**, 42-45 (1954).

[16] D. A. S. Frazer. Canadian J. Math. **6**, 46—68 (1954).

[17] D. Basu. Ann. math. Statistics **26**, 345—348 (1955).

[18] H. Cramér. Skand. Aktuarietidskr. **29**, 85—94 (1946).

[19] D. A. S. Frazer und I. Guttmann. Ann. math. Statistics **27**, 162—179 (1956).

[20] D. A. S. Frazer. Ann. math. Statistics **24**, 44—55 (1953). Man vgl. auch den Bericht von S. S. Wilks. Bull. Amer. math. Soc. **54**, 6—50 (1948).

[21] Vgl. J. Neyman und E. S. Pearson. Biometrika **20** A, 175—240; 263—294 (1928), Philos. Trans. Roy. Soc. **231**, 289—337 (1933) u. a.

[22] J. Neyman und E. S. Pearson. Statistical Research Mem. **1**, 1—37 (1936).

[23] G. B. Dantzig und A. Wald. Ann. math. Statistics **22**, 87—93 (1951). Dort orientiere man sich auch für eine präzise Fassung der Eindeutigkeitsaussage.

[24] A. Wald. Trans. Amer. math. Soc. **54**, 462—482 (1943).