Werk

Titel: Kapazitäten statt Wahrscheinlichkeiten? Gedanken zur Grundlegung der Statistik.

Autor: Huber, Peter J.

PURL: https://resolver.sub.uni-goettingen.de/purl?37721857X_0078|log9

Kontakt/Contact

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

info@digizeitschriften.de

Jber. d. Dt. Math.-Verein
78 (1976) 81 – 92

Kapazitäten statt Wahrscheinlichkeiten? Gedanken zur Grundlegung der Statistik

Von PETER J. HUBER in Zürich

1. Einleitung. Wenn ein angewandter Statistiker seinen Beobachtungen eine "Normalverteilung" zuschreibt, so ist das nicht allzu wörtlich zu nehmen: seine Meßwerte werden auf wenige Dezimalstellen gerundet sein und in einem von vornherein beschränkten Meßbereich liegen; außerdem wird man mit noch weiteren, nicht auf diese Diskretisierungseffekte zurückzuführenden Abweichungen zu rechnen haben.

Man pflegt derartige Diskrepanzen unter Anrufung eines unbestimmten Stetigkeitsprinzips zu vernachlässigen. Das kann gefährlich sein und ist es im vorliegenden Fall tatsächlich auch: die herkömmlichen, für die Normalverteilung optimierten Methoden sind für statistisch sinnvolle Topologien nämlich nirgends stetig.

Als statistisch sinnvoll ist insbesondere die schwache Topologie zu betrachten. Dies soll kurz plausibel gemacht werden. Eine Mehrheit der statistischen Verfahren, sowohl Teststatistiken wie Schätzer, hängt von den Beobachtungen $x_1, \ldots, x_n$ nur über die empirische Verteilung

$$\mu_n = \frac{1}{n} \Sigma \delta_{x_i}$$

ab, wo δ_x die Punktmasse 1 in x bezeichnet, d. h., sie können geschrieben werden als $T_n(x_1, \ldots, x_n) = T(\mu_n)$.

Beispiele. (1) Der arithmetische Mittelwert $\bar{x} = \int x \, d\mu_n$. (2) Die Maximum-Likelihood-Schätzung des Parameters θ in einer Familie $f(x, \theta)$ von Dichten ist dadurch definiert, daß sie $\int f(x, \theta) d\mu_n$ maximiert.

Wenn T speziell ein lineares Funktional ist, erhalten wir

$$T(\mu_n) = \frac{1}{n} \sum T(\delta_{x_i}) = \int \psi \, d\mu_n$$

mit $\psi(x) = T(\delta_x)$. Das anfänglich angerufene Stetigkeitsprinzip besagt nun doch wohl, daß $T(\mu_n)$ bei kleinen Änderungen aller Beobachtungen („Rundungsfehler") sowie bei großen Änderungen einiger weniger Beobachtungen

(„grobe Fehler") sich nur wenig ändern soll. Daraus folgt, daß ψ stetig und beschränkt sein soll, und das führt in natürlicher Weise zur allgemeinen Forderung, daß die statistischen Funktionale T (lineare und nichtlineare) in bezug auf die schwache Topologie in der Menge $\mathcal{M}_1$ der Wahrscheinlichkeitsmaße auf R stetig sein sollen, d. h. in bezug auf die schwächste, für welche alle Abbildungen

$$\mu \to \int \psi \, d\mu$$

(mit stetigem und beschränktem ψ) stetig sind. Die traditionellen statistischen Verfahren, wie zum Beispiel der arithmetische Mittelwert $\bar{x} = \int x \, d\mu_n$, sind leider nicht stetig in diesem Sinne, oder, wie man sich heute auszudrücken pflegt, nicht „robust". Die Theorie der qualitativen Robustheit ist im wesentlichen eine Theorie der schwach stetigen (nicht-linearen) Funktionale auf $\mathcal{M}_1$ (vgl. Hampel 1971).

Eine analoge und ebenfalls ungemütlich große Kluft besteht zwischen dem numerischen Rechnen und dem platonischen Ideal der reellen Zahlen. Dort ist man sich des Konflikts schon längere Zeit bewußt, und man überbrückt ihn etwa so, daß man das Rechnen im endlichen Bereich der Gleitkommazahlen axiomatisiert und dann diesen mit dem idealen Bereich der reellen Zahlen durch die Einführung von Intervallarithmetik verknüpft.

Ich möchte nun ein dadurch inspiriertes Programm für den Fall der Statistik skizzieren. Es geht darum, eine Beschreibung von approximativ gegebenen Wahrscheinlichkeiten nicht nur qualitativ-topologisch, sondern durch direkte Angabe von unteren und oberen Schranken zu formulieren. Wenn möglich, sollte diese Beschreibung so beschaffen sein, daß sie ungeändert von den endlichen Stichprobenräumen der Praxis auf die herkömmlichen allgemeineren Räume der Wahrscheinlichkeitstheorie (mindestens die euklidischen) übertragen werden kann.

In der Literatur sind mehrere Ansätze zu einer Art Intervallarithmetik für Wahrscheinlichkeiten anzutreffen. Der wohl erste stammt von G. Boole (1854, Kap. XVII ff.). Er nimmt an, daß nur für eine gewisse Kollektion von Ereignissen Wahrscheinlichkeiten bekannt seien, und er berechnet daraus Schranken für die Wahrscheinlichkeit eines beliebigen Ereignisses.

Beispiel: $P(A) = p$ und $P(A \cap B) = q$ seien bekannt, dann gilt $q \leq P(B) \leq q + 1 - p$.

Koopmans (1940) axiomatisches Modell für die subjektive Wahrscheinlichkeitsauffassung führt zunächst auf partiell geordnete, nicht-numerische Wahrscheinlichkeiten. Der Vergleich mit sogenannten „Wahrscheinlichkeitsskalen" (n gleichwahrscheinliche, sich gegenseitig ausschließende Aussagen) erlaubt dann, für jedes beliebige A eine untere und eine obere numerische Wahrscheinlichkeit $P_*(A)$, respektive $P^*(A)$, einzuführen.

Koopman beschränkt sich dann etwas voreilig auf die Aussagen, die „bewertbar" (appraisable) sind: $P_*(A) = P^*(A)$.

C. A. B. Smith (1961) weist darauf hin, daß die gängige Begründung des persönlichen Wahrscheinlichkeitsbegriffs durch eine Formalisierung des konsistenten Wettverhaltens nicht auf eindeutig bestimmte Wahrscheinlichkeitswerte, sondern nur auf Schranken $P_* \leq P^*$ führt (zwischen welchen Wettangeboten man weder die eine noch andere Seite zu ergreifen gewillt ist).

Weitere Ansätze finden sich bei I. J. Good (1962), Strassen (1964), der als erster den Zusammenhang mit Choquets Kapazitäten (1953/54) erkannt und ausgenützt hat, Dempster (1968) und Beran (1971). Die Modellvorstellungen und die Forderungen, die von den verschiedenen Autoren an die Mengenfunktionen P_*, P^* gestellt werden, sind durchaus verschieden; es soll nun eine formale Klärung und Straffung der verschiedenen Ansätze versucht werden.

Es erscheint zweckmäßig, vier Stufen von abnehmender Allgemeinheit herauszusondern und getrennt zu betrachten. Im folgenden bezeichnet $\mathscr{P}$ eine Menge von Wahrscheinlichkeitsmaßen auf einer (vorerst endlichen) Menge Ω.

(0) Nichtlineare Funktionale $\tilde{u}, \tilde{v}$, die als scharfe Schranken einer Kollektion von Erwartungen darstellbar sind:

$$\tilde{u}(f) = \inf\{\int f \, d\mu | \mu \in \mathscr{P}\}$$
$$\tilde{v}(f) = \sup\{\int f \, d\mu | \mu \in \mathscr{P}\}.$$

(I) Nichtadditive Mengenfunktionen u, v, die als scharfe Schranken einer Kollektion von Wahrscheinlichkeitsmaßen darstellbar sind:

$$u(A) = \inf\{\mu(A) | \mu \in \mathscr{P}\}$$
$$v(A) = \sup\{\mu(A) | \mu \in \mathscr{P}\}.$$

(II) 2-monotone, respektive 2-alternierende Mengenfunktionen u, respektive v.
(III) Totalmonotone, respektive totalalternierende Mengenfunktionen u, respektive v.

Weil die unteren und oberen Schranken jeweils durch

$$\tilde{v}(f) = -\tilde{u}(-f),$$
$$v(A) = 1 - u(A^c)$$

verknüpft sind, genügt es in der Regel, nur die eine zu betrachten.

Stufe (0) ist vollkommen überblickbar. Ein Funktional $\tilde{v}$ ist genau dann als scharfe obere Schranke einer Menge $\mathscr{P}$ darstellbar, wenn $\tilde{v}$ monoton:

$$f \leq g \Rightarrow \tilde{v}(f) \leq \tilde{v}(g),$$

positiv affin homogen:

$$\tilde{v}(af + b) = a\tilde{v}(f) + b, \quad a,b \in \mathbb{R}, a \geq 0,$$

und subadditiv ist:

$$\tilde{v}(f + g) \leq \tilde{v}(f) + \tilde{v}(g).$$

Außerdem besteht eine natürliche eineindeutige Entsprechung zwischen den konvexen, abgeschlossenen Mengen $\mathscr{P}$ und ihren Stützfunktionen $\tilde{v}$. Für die Verallgemeinerung auf unendliche Räume vgl. Hörmander (1954). Da wir Wahrscheinlichkeitsmaße primär als Mengenfunktionen betrachten wollen, lassen wir diese Stufe im folgenden beiseite.

2. Untere und obere Schranken für Wahrscheinlichkeiten auf endlichen Räumen. Es geht hier um notwendige und hinreichende Bedingungen, die eine Mengenfunktion erfüllen muß, damit sie eine scharfe untere/obere Schranke für eine gewisse Kollektion $\mathscr{P}$ von Wahrscheinlichkeitsmaßen auf einem endlichen Raum Ω darstellt.

Die offensichtlichen notwendigen Bedingungen:

- u, v monoton zunehmend,
- $v(A) = 1 - u(A^c)$
- $u(\emptyset) = v(\emptyset) = 0, \quad u(\Omega) = v(\Omega) = 1$
- $u(A \cup B) \geq u(A) + u(B)$ für $A \cap B = \emptyset$
- $v(A \cup B) \leq v(A) + v(B)$

sind nicht hinreichend.

Beispiel (G. Wolf). Sei $|\Omega| = 4$, und $u(A), v(A)$ hängen nur von der Kardinalität der Menge A ab, gemäß der folgenden Tafel.

$\|A\|$	0	1	2	3	4
u	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1
v	0	$\frac{1}{2}$	$\frac{1}{2}$	1	1

Hier gibt es nur eine einzige additive Mengenfunktion μ zwischen u und v, nämlich $\mu(A) = |A|/4$.

Sei $\mathscr{M}_1$ die Menge aller Wahrscheinlichkeitsmaße auf Ω, sei $\mathscr{D}$ eine Kollektion von Teilmengen von Ω, und sei $u: \mathscr{D} \to \mathbb{R}_+$ eine beliebige nichtnegative Mengenfunktion. Wir bezeichnen mit

$$\mathscr{P} = \{\mu \in \mathscr{M}_1 | \mu(A) \geq u(A), \ \forall A \in \mathscr{D}\}$$

die Menge der Wahrscheinlichkeitsmaße „oberhalb" von u. Dual können wir $\mathscr{P}$ charakterisieren als die Menge

$$\mathscr{P} = \{\mu \in \mathscr{M}_1 | \mu(B) \leq v(B),\ \forall B \text{ mit } B^c \in \mathscr{D}\},$$

mit $v(B) = 1 - u(B^c)$.

Satz 2.1. *$\mathscr{P}$ ist genau dann nicht leer, wenn aus*

$$\sum a_i 1_{A_i} \leq 1 \quad (a_i \geq 0,\ A_i \in \mathscr{D})$$

folgt

$$\sum a_i u(A_i) \leq 1\,.$$

Beweis. Die Notwendigkeit der Bedingung ist klar. Daß sie hinreichend ist, folgt aus dem nächsten Satz 2.2.

Für reelle Funktionen $f : \Omega \to \mathbb{R}$ definieren wir die *untere Erwartung*

$$\tilde{u}(f) = \sup\{\textstyle\sum a_i u(A_i) - a \,|\, \sum a_i 1_{A_i} - a \leq f,\ a_i \geq 0,\ A_i \in \mathscr{D}\}$$

sowie die *untere Wahrscheinlichkeit*

$$u_0(A) = \tilde{u}(1_A),\quad A \subset \Omega\,.$$

Dual erhalten wir eine obere Erwartung

$$\tilde{v}(f) = -\tilde{u}(-f) = \inf\{\textstyle\sum b_i v(B_i) - b \,|\, \sum b_i 1_{B_i} - b \geq f,\ b_i \geq 0,\ B_i^c \in \mathscr{D}\}\,,$$

und eine obere Wahrscheinlichkeit

$$v_0(A) = \tilde{v}(1_A),\quad A \subset \Omega\,.$$

Man überlegt sich leicht, daß

$$u(A) \leq u_0(A) \quad \text{für} \quad A \in \mathscr{D}$$

und daß aus $\mu \geq u$ folgt $\mu \geq u_0$ sowie, daß die Wiederholung der ganzen Konstruktion nichts Neues liefert:

$$(u_0)^{\sim} = \tilde{u}\,,$$

und speziell

$$u_{00} = u_0\,.$$

Duales gilt für die von v abgeleiteten Größen.

Satz 2.2. *Für alle $f : \Omega \to \mathbb{R}$ gilt*

$$\begin{aligned}\tilde{u}(f) &= \inf\{\textstyle\int f \,\mathrm{d}\mu | \mu \in \mathscr{P}\}\\ \tilde{v}(f) &= \sup\{\textstyle\int f \,\mathrm{d}\mu | \mu \in \mathscr{P}\}\end{aligned}$$

und insbesondere

$$\begin{aligned}u_0(A) &= \inf\{\mu(A) | \mu \in \mathscr{P}\}\\ v_0(A) &= \sup\{\mu(A) | \mu \in \mathscr{P}\}\,.\end{aligned}$$

Wenn $\mathscr{P}$ leer ist, sind $\tilde{u}$ und u_0 identisch $+\infty$; andernfalls sind $\tilde{u}, u_0$ überall endlich.

Beweis. Man zeigt zuerst, daß $\tilde{u}(f) \geq 0$ falls $f \geq 0$ und daß entweder $\tilde{u}(0) = 0$ oder aber $\tilde{u}(f) = \infty$ für alle f. Im zweiten Fall ist $\mathscr{P} = \emptyset$. Im ersten Fall verifiziert man leicht, daß $\tilde{u}, \tilde{v}$ monoton, positiv affin homogen und superadditiv (respektive subadditiv) sind. Die Behauptung folgt nun leicht mit Hilfe des Satzes von Hahn-Banach; vergleiche auch den Schluß des vorangegangenen Abschnitts. Satz 2.1 folgt nun ebenfalls, denn die dort gemachte Voraussetzung ist mit $\tilde{u}(0) \leq 0$ gleichbedeutend.

Sei jetzt $\mathscr{D} = 2^{\Omega}$. Dann ist $u = u_0$ eine offensichtliche notwendige und hinreichende Bedingung dafür, daß u eine scharfe untere Schranke für eine Menge von Wahrscheinlichkeiten ist. Wenn wir diese Bedingung ausschreiben, kommt das

Korollar 2.3. *Notwendig und hinreichend für*

$$u(A) = \inf\{\mu(A) | \mu \in \mathscr{M}_1, \mu \geq u\}, A \subset \Omega,$$

ist, daß aus

$$1_A \geq \sum a_i 1_{A_i} - a, \quad (a_i \geq 0, A_i \subset \Omega),$$

folgt

$$u(A) \geq \sum a_i u(A_i) - a.$$

Man beachte, daß im allgemeinen mehrere verschiedene superadditive Funktionale $\tilde{u}'$ dieselbe Mengenfunktion $u_0(A) = \tilde{u}'(1_A)$ induzieren; $\tilde{u}$ ist unter ihnen das kleinste.

3. Unendliche Räume. Wir benützen im folgenden wesentlich, daß „genügend viele" reguläre Maße vorhanden sind. Der Einfachheit halber beschränken wir uns auf lokalkompakte Räume Ω mit abzählbarer Basis; vieles läßt sich ungeändert auf polnische Räume (mit den Borel-Maßen) oder auf allgemeine lokalkompakte Räume (mit den Baire- bzw. Radon-Maßen) übertragen.

Eine Menge P von Wahrscheinlichkeitsmaßen auf Ω soll durch Intervallangaben $u(A) \leq \mu(A) \leq v(A)$, gültig für alle Borelmengen A und alle $\mu \in \mathscr{P}$, beschrieben werden. Aufgrund eines (vorläufig noch etwas unbestimmten) Stetigkeitsprinzips verlangen wir:

(i) In Analogie zur bekannten Regularitätseigenschaft für Wahrscheinlichkeitsmaße

$$\mu(A) = \sup\{\mu(K) | K \subset A, K \text{ kompakt}\}$$

für alle Borelmengen A, sollen auch u und v durch ihre Werte auf kompakten Mengen eindeutig bestimmt sein.

(ii) Wenn alle $\mu \in \mathscr{P}$ die Ungleichungen $u \leq \mu \leq v$ erfüllen, dann soll dies auch für die abgeschlossene Hülle $\bar{\mathscr{P}}$ (in einer geeigneten Topologie) gelten.

Um (i) zu erzwingen, definieren wir u, v auf der Klasse $\mathscr{K}$ der kompakten Teilmengen von Ω und erweitern dann auf die Klasse $\mathscr{A}$ der Borelmengen.

Satz 3.1. *Sei $\mathscr{P}$ eine beliebige nicht-leere Menge von Wahrscheinlichkeitsmaßen auf (Ω, A). Dann definiert*

$$u(K) = \inf\{\mu(K) | \mu \in \mathscr{P}\}, \quad K \in \mathscr{K},$$

eine Kapazität (im ursprünglichen Sinn von Choquet 1953/54) auf der Klasse $\mathscr{K}$, d. h., u ist monoton wachsend und auf $\mathscr{K}$ von rechts stetig.

Wir erinnern an die Definition der Stetigkeit von rechts in $A \in \mathscr{K}$ (Choquet 1953/54, p. 174): für jede Umgebung W von $u(A)$ existiert eine Umgebung V von A, derart, daß für alle $X \in \mathscr{K}$ mit $A \subset X \subset V$ gilt $u(X) \in W$.

Der Beweis dieses Satzes benützt nur, daß alle $\mu \in \mathscr{P}$ regulär sind und ergibt sich mit Leichtigkeit aus den folgenden Definitionen und Bemerkungen.

Zunächst definieren wir drei Mengenfunktionen u_0, u_*, u^*:

1. Die untere Wahrscheinlichkeit

$$u_0(A) = \inf\{\mu(A) | \mu \in \mathscr{P}\} \quad \text{für} \quad A \in \mathscr{A}.$$

2. Die innere Kapazität

$$u_*(A) = \sup\{u_0(K) | K \subset A, K \in \mathscr{K}\} \quad \text{für} \quad A \subset \Omega.$$

3. Die äußere Kapazität

$$u^*(A) = \inf\{u_*(G) | G \supset A, G \text{ offen}\} \quad \text{für} \quad A \subset \Omega.$$

Alle diese Mengenfunktionen sind monoton, und aus den Definitionen folgt unmittelbar, daß $u_*(A) \leq u^*(A)$ für beliebige $A \subset \Omega$ sowie daß $u_*(A) \leq u_0(A)$ für $A \in \mathscr{A}$. Außerdem haben wir $u^*(A) = \inf_{G \supset A} u_*(G) \leq \inf_{G \supset A} u_0(G) = \inf_G \inf_\mu \mu(G) = \inf_\mu \inf_G \mu(G) = \inf_\mu \mu(A) = u_0(A)$. Somit gilt $u^*(A) \leq u_0(A)$ für $A \in \mathscr{A}$. Für kompaktes K haben wir $u_*(K) = u_0(K)$, und somit auch $u^*(K) = u_0(K)$, was die Stetigkeit von rechts ergibt.

Eine Menge $A \subset \Omega$ heißt *u-kapazitabel*, wenn $u_*(A) = u^*(A)$, und wir definieren für beliebige kapazitable Mengen

$$u(A) = u_*(A) = u^*(A).$$

Insbesondere sind die kompakten und die offenen Mengen u-kapazitabel.

Satz 3.2 (Choquet). *Die zu u konjugierte Mengenfunktion v, definiert durch* $v(K) = 1 - u_*(K^c)$, *ist ebenfalls eine Kapazität auf* $\mathscr{K}$.

Analog definiert man nun innere und äußere Kapazitäten und eine obere Wahrscheinlichkeit durch

$$\begin{aligned} v_*(A) &= \sup\{v(K) | K \subset A, K \in X\} && \text{für} \quad A \subset \Omega, \\ v^*(A) &= \inf\{v_*(G) | G \supset A, G \text{ offen}\} && \text{für} \quad A \subset \Omega, \\ v_0(A) &= \sup\{\mu(A) | \mu \in \mathscr{P}\} && \text{für} \quad A \in \mathscr{P}. \end{aligned}$$

Satz 3.3. *Die nachstehenden Ungleichungen gelten für alle* $A \in \mathscr{A}$ *(die äußersten für beliebige* $A \subset \Omega$*):*

$$u_*(A) \leq u^*(A) \leq u_0(A) \leq v_0(A) \leq v_*(A) \leq v^*(A).$$

Bemerkung. Für $A \notin \mathscr{A}$ gilt $u^*(A) \leq v_*(A)$ im allgemeinen nicht. Gegenbeispiel: $\mathscr{P} = \{\mu\}$, A nicht μ-meßbar.

Sei $\mathscr{P}$ eine beliebige nichtleere Menge von Wahrscheinlichkeitsmaßen auf Ω, und sei $\bar{\mathscr{P}}$ die vage Abschließung von $\mathscr{P}$ in der Menge der substochastischen Maße. (Die vage Topologie ist die schwächste, für welche $\mu \to f \mathrm{d}\mu$ für alle stetigen f mit kompaktem Träger stetig ist.) $\bar{\mathscr{P}}$ ist vag kompakt.

Satz 3.4. $\mathscr{P}$ *und* $\bar{\mathscr{P}}$ *erzeugen die gleichen Kapazitäten u,v.*

Beweis. Man verifiziert leicht, daß aus $\mu_n \to \mu$ für kompaktes K folgt $\limsup \mu_n(K) \leq \mu(K)$.

Wenn $\mathscr{P}$ eine endliche Menge ist, haben wir in Satz 3.3 überall (außer in der Mitte) Gleichheitszeichen, für beliebige Borelmengen A.

Allgemein gilt: wenn $\mathscr{P}$ schwach kompakt ist, haben wir $u^*(A) = u_0(A)$ und $v_0(A) = v_*(A)$ für $A \in \mathscr{A}$. Man beachte in diesem Zusammenhang, daß $\mathscr{P}$ genau dann straff ist, wenn $u_*(\Omega) = 1$ bzw. $v(\emptyset) = 0$.

Ich vermute, daß für alle Borelmengen A auch $u_*(A) = u^*(A)$ und $v_*(A) = v^*(A)$ gilt, sofern $\mathscr{P}$ schwach kompakt ist; bewiesen ist diese Aussage aber nur für 2-monotones u respektive 2-alternierendes v (die gegenteilige Behauptung in Huber und Strassen (1973) beruht auf einer irrtümlichen Vermischung von gewöhnlicher und abstrakter Kapazitabilität, Choquet (1953/54) bzw. (1959)).

4. Totalmonotone und totalalternierende Kapazitäten. Wir schieben die Behandlung von Stufe (II) noch auf und gehen gleich zur eingeschränktesten dritten Stufe über. Diese erhalten wir durch eine direkte Probabilisierung der Intervallarithmetik (Strassen 1964), und zwar machen wir uns folgende Modellvorstellung. Das Zufallsexperiment besteht nach wie vor aus dem zufälligen Herausgreifen eines Elementarereignisses ω, liefert aber nicht

mehr einen einzigen Meßwert, sondern eine Menge $T(\omega) \subset \mathbb{R}$ von möglichen Meßwerten, von denen der Statistiker jeweils nur einen, irgendwie herausgegriffenen, zu sehen bekommt. Sei nun $A \subset \mathbb{R}$, dann sind

$$E_* = \{\omega | T(\omega) \subset A\}$$

und

$$E^* = \{\omega | T(\omega) \cap A \neq \emptyset\}$$

die Mengen der Elementarereignisse, welche einen Wert in A liefern müssen bzw. können. Die Wahrscheinlichkeit des Eintreffens von A wird also durch

$$u(A) = P(E_*)$$

und

$$v(A) = P(E^*)$$

nach unten bzw. oben beschränkt.

Abgesehen von gewissen Regularitäts- und Stetigkeitsbedingungen läßt sich eine Mengenfunktion u (bzw. v) genau dann in der obigen Art mittels eines mengenwertigen stochastischen Prozesses darstellen, wenn sie $u(\emptyset) = 0$ und $u(\mathbb{R}) = 1$ erfüllt und totalmonoton (bzw. totalalternierend) ist (Satz von Choquet 1953/54, p. 270ff.).

Wir erinnern an die Definition von monotonen und alternierenden Mengenfunktionen ψ. Seien $X, A_1, \ldots, A_p$ Teilmengen von Ω, und seien

$$\Lambda_p = \psi(X) - \sum_i \psi(X \cap A_i) + \sum_{i<j} \psi(X \cap A_i \cap A_j) + \cdots + (-1)^p \psi(X \cap A_1 \cap A_2 \cap \cdots \cap A_p)$$

$$V_p = \psi(X) - \sum_i \psi(X \cup A_i) + \sum_{i<j} \psi(X \cup A_i \cup A_j) + \cdots + (-1)^p \psi(X \cup A_1 \cup A_2 \cup \cdots \cup A_p).$$

Sei $n \geq 2$. Wenn für alle $p \leq n$, alle X und alle p-tupel $A_1, \ldots, A_p$ gilt $\Lambda_p \geq 0$, dann heißt ψ *n-monoton*, wenn dual $V_p \leq 0$, dann heißt ψ *n-alternierend*. Falls diese Beziehungen für alle n gelten, spricht man von *totalmonotonen*, respektive *totalalternierenden* Mengenfunktionen.

Dieser Ansatz trägt weiter, als es zunächst scheinen möchte. Seien zum Beispiel Y, U zwei unabhängige Zufallsgrößen, die erste mit der idealisierten Verteilung P_0, und die zweite nehme die zwei Werte $\delta \geq 0$ und $+\infty$ mit den Wahrscheinlichkeiten $1 - \varepsilon$ und ε an. Setze $T(\omega) = [Y(\omega) - U(\omega), Y(\omega) + U(\omega)]$. Bei jeder Durchführung des Experiments bekommt der Statistiker eine irgendwie (nicht unbedingt zufällig) ausgewählte Zahl $x \in T(\omega)$ zu sehen. Mit Wahrscheinlichkeit $1 - \varepsilon$ ist also x nur mit einem

kleinen Rundungsfehler $\leq \delta$ behaftet, mit Wahrscheinlichkeit ε aber mit einem beliebig großen groben Fehler.

Das Hauptziel der robusten Statistik ist, gegen derartige Fehler (vor allem gegen die gelegentlichen groben Fehler) unempfindliche Methoden zu schaffen. Das führt auf Minimaxprobleme: man möchte sich gegen den schlimmstmöglichen Fall unter den Wahrscheinlichkeiten zwischen u und v absichern. Nun stellte sich heraus, daß einerseits die wichtigsten Sätze bereits für 2-monotone/2-alternierende Mengenfunktionen u/v gelten und daß andererseits die in der robusten Statistik auftretenden u, v, wenn wir von den oben konstruierten absehen, meist zwar 2-monoton/2-alternierend, nicht aber totalmonoton/totalalternierend sind. Dies, und die Bemerkung am Schluß von Abschnitt 3, liefern eine starke Motivation zur gesonderten Betrachtung von 2-monotonen/2-alternierenden Kapazitäten.

Zu den totalmonotonen/totalalternierenden Kapazitäten ist beizufügen, daß sie in den letzten Jahren im Zusammenhang mit den geometrischen Wahrscheinlichkeiten, d. h. mengenwertigen stochastischen Prozessen, Wichtigkeit erlangt haben und von Kendall (1974) und Matheron (1975) ausführlich behandelt worden sind.

5. 2-monotone und 2-alternierende Kapazitäten. Mengenfunktionen u (bzw. v) sind genau dann 2-monoton (2-alternierend), wenn sie monoton wachsend sind und

$$u(A \cup B) + u(A \cap B) \geq u(A) + u(B)$$

bzw.

$$v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$$

erfüllen.

Für unsere Zwecke entscheidend ist die Äquivalenz dieser Eigenschaften mit der folgenden: für jede monotone Folge $A_1 \subset A_2 \subset \cdots \subset A_n$ existiert ein Maß $\mu \leq v$, welches die Wahrscheinlichkeit von A_i für alle i *gleichzeitig* maximiert: $\mu(A_i) = v(A_i)$. Duales gilt für u.

Diese Eigenschaft läßt sich auch so ausdrücken: für $f \geq 0$ gilt

$$\tilde{v}(f) = \sup\{\int f \mathrm{d}\mu | \mu \leq v\} = \int_0^\infty v\{x | f(x) > t\} \mathrm{d}t$$

(vgl. Satz 2.2). Für die Regularitätsbedingungen, die im Fall unendlicher Räume zu stellen sind, siehe Huber und Strassen (1973).

Derartige Mengenfunktionen treten in natürlicher Weise bei der Formalisierung von Problemen der robusten Statistik auf.

Beispiele: Sei P_0 ein festes Wahrscheinlichkeitsmaß auf einem kompakten Raum Ω.

(1) Setze $v_0(A) = \min[P_0(A) + \varepsilon, 1]$ für kompaktes $A \neq \emptyset$. Dann ist $P \leq v_0$ äquivalent mit $\|P - P_0\| \leq \varepsilon$ ($\| \|$: totale Variation).
(2) Setze $v_0(A) = \min[P_0(A^\varepsilon) + \varepsilon, 1]$ für kompaktes $A \neq \emptyset$, wobei A^ε die abgeschlossene ε-Umgebung von A bezeichnet. Dann ist $P \leq v_0$ äquivalent mit $d_{Pr}(P_0, P) \leq \varepsilon$ (d_{Pr} Prohorov-Abstand).
In beiden Fällen ist v_0 2-alternierend.

Der bisher tiefste Satz mit statistischer Relevanz ist das Neyman-Pearson-Lemma für Kapazitäten von Strassen (1964). Ursprünglich für totalmonotone/totalalternierende Kapazitäten auf endlichen Räumen formuliert, läßt es sich jedoch verallgemeinern für 2-monotone/2-alternierende Kapazitäten auf polnischen Räumen (Huber und Strassen 1973).

Satz 5.1 (Neyman-Pearson-Lemma für Kapazitäten). *Seien* $\mathcal{P}_0 = \{P \leq v_0\}$ *und* $\mathcal{P}_1 = \{P \leq v_1\}$ *zwei disjunkte Mengen von Wahrscheinlichkeitsmaßen, die sich durch 2-alternierende Kapazitäten* v_0, v_1 *definieren lassen. Dann existiert ein „ungünstigstes Paar"* (Q_0, Q_1), $Q_i \in \mathcal{P}_i$, *und eine Version* $\pi = \mathrm{d}Q_1/\mathrm{d}Q_0$ *der Radon-Nikodym-Ableitung, derart, daß für alle* $t \in \mathbb{R}$ *und alle* $Q_i' \in \mathcal{P}_i$ *gilt*

$$Q_0\{\pi > t\} \geq Q_0'(\pi > t\}, \quad Q_1\{\pi > t\} \leq Q_1'\{\pi > t\}.$$

Bemerkung. Die Funktion π (nicht aber das Paar Q_0, Q_1) ist durch v_0, v_1 im wesentlichen eindeutig bestimmt und kann als verallgemeinerte Radon-Nikodym-Ableitung $\mathrm{d}v_1/\mathrm{d}v_0$ aufgefaßt werden.

Korollar 5.2. *Für alle Stichprobengrößen n und alle Niveaus* α *existiert ein Minimax-Test zwischen* $\mathcal{P}_0$ *und* $\mathcal{P}_1$. *Dieser verwirft bzw. akzeptiert* $\mathcal{P}_0$, *wenn*

$$\prod_{i=1}^{n} \pi(x_i) > C \quad \text{bzw.} \quad < C,$$

für eine gewisse Konstante C.

Bemerkung. Das Überraschende an diesem Resultat ist, daß dasselbe ungünstigste Paar (Q_0, Q_1) bzw. dasselbe $\pi(x)$ für alle α und n gültig ist.

Wir haben hier nun an der Oberfläche eines großen Problemkreises gekratzt, und die Verzweigungen werden erst allmählich sichtbar. Zum Beispiel lassen sich aus dem Neyman-Pearson-Lemma nicht nur robuste Tests, sondern auch robuste Schätzungen gewinnen (Huber 1968). Die ungünstigsten Paare (Q_0, Q_1) lassen sich durch ε-Suffizienz charakterisieren, was eine Querverbindung zur Theorie des Vergleichs von Experimenten ergibt (Le Cam 1965, Torgersen 1970, Buja 1975). Kapazitäten scheinen eine natürliche Formalisierung der Ungenauigkeit des a-priori-Wissens zu

ermöglichen, welche den unendlichen Regreß (a-priori-Verteilungen auf der Menge der a-priori-Verteilungen usw.) des reinen Bayes-Ansatzes durchbricht.

Literaturverzeichnis

Beran, R. J.: On distribution-free statistical inference with upper and lower probabilities. Ann. Math. Statist. **42** (1971) 157–168.

Boole, G.: An investigation of the laws of thought. 1854 (Nachdr. New York 1958).

Buja, A.: Diplomarbeit ETH Zürich 1975.

Choquet, G.: Theory of capacities. Ann. Inst. Fourier **5** (1953/54) 131–295.

Choquet, G.: Forme abstraite du théorème de capacitabilité. Ann. Inst. Fourier **9** (1959) 83–89.

Dempster, A. P.: A generalization of Bayesian inference. J. Royal Statist. Soc. **B 30** (1968) 205–245.

Good, I. J.: The measure of a non-measurable set. In: Logic Methodology and Philosophy of Science (ed. E. Nagel, P. Suppes, A. Tarski). Stanford, Calif. 1962, p. 319–329.

Hampel, F. R.: A general qualitative definition of robustness. Ann. Math. Statist. **42** (1971) 1887–1896.

Hörmander, L.: Sur la fonction d'appui des ensembles convexes dans un espace localement convexe. Arkiv. Mat. **3** (1954) 181–186.

Huber, P. J.: Robust confidence limits. Z. Wahrscheinlichkeitstheorie **10** (1968) 269–278.

Huber, P. J.; Strassen, V.: Minimax tests and the Neyman-Pearson Lemma for capacities. Ann. Statist. **1** (1973) 251–263 und **2** (1974) 223–224.

Kendall, D. G.: Foundations of a Theory of Random Sets. In: Stochastic Geometry (ed. E. F. Harding and D. G. Kendall), London-New York-Sidney-Toronto 1974, p. 322–376.

Le Cam, L.: Sufficiency and approximate sufficiency. Ann. Math. Statist. **35** (1965) 1419–1455.

Koopman, B. O.: The axioms and algebra of intuitive probability. Ann. Math. **41** (1940) 269–292.

Matheron, G.: Random sets and integral geometry. London-New York-Sidney-Toronto 1975.

Smith, C. A. B.: Consistency in statistical inference and decision. J. Royal Statist. Soc. **B 23** (1961) 1–25.

Strassen, V.: Meßfehler und Information. Z. Wahrscheinlichkeitstheorie **2** (1964) 273–305.

Torgersen, E. N.: Comparison of experiments when the parameter space is finite. Z. Wahrscheinlichkeitstheorie **16** (1970) 219–249.

(Eingegangen: 7.11.1975)

Eidgenössische Technische Hochschule
Zürich
CH 8006 Zürich
Claussiusstr. 55