Probab. Th. Rel. Fields 73, 149–152 (1986)

Probability Theory and Related Fields

On the Huber-Strassen Theorem*

Andreas Buja

Department of Statistics, GN-22, University of Washington, Seattle, WA 98195, USA, and
Bell Communications Research, 2M-335, 435 South Street Morristown, NJ 07960, USA

Summary. We show that the Huber-Strassen theorem can be applied to many common robust neighborhood models without restrictive compactness assumptions. Our method is to replace the original Polish topology with a compact topology, which is possible according to the Kuratowski isomorphism theorem.

1. Introduction

In the robustness literature, it is common practice to replace exact probability models by "approximate models", formalized by some sort of convex neighborhoods or sets $\mathbf{Q}$ of probability measures. Most frequently, we encounter "contamination neighborhoods" given by

$$\begin{aligned}\mathbf{Q} &= \{(1-\varepsilon)\cdot P + \varepsilon\cdot H \mid H = \text{arbitrary probability measure}\}\\ &= \{Q \text{ probability measure} \mid Q(A) \leqq (1-\varepsilon)\cdot P(A) + \varepsilon\ \forall A \text{ measurable}\},\end{aligned}$$

which have been used in the theory both of robust estimation [5] and robust testing [6]. Subsequently, Huber [7], and after him Rieder [10], have gone on to a more flexible family of sets which comprise contamination as well as neighborhoods in variation norm:

$$\mathbf{Q} = \{Q \text{ probability measure} \mid Q(A) \leqq (1-\varepsilon)\cdot P(A) + \varepsilon + \delta\ \forall A \text{ measurable}\}.$$

The upper probability

$$w(A) = \sup_{Q\in\mathbf{Q}} Q(A) = ((1-\varepsilon)\cdot P(A) + \varepsilon + \delta)\wedge 1 \qquad \text{if } A \neq \emptyset,\ w(\emptyset) = 0,$$

completely describes the neighborhood $\mathbf{Q}$ and is analytically more tractable as a set function. In the derivation of their robust Neyman-Pearson lemma,

* This work was supported by the following grants: *NSF*, MCS-8304234 and DMS-8504359; *ONR*, N00014-81-k-0095; *DoE*, DE-FG06-85ER25006

Huber and Strassen [8], therefore, worked with set functions which have all the properties of upper probabilities over suitably chosen convex sets of probability measures, namely so called 2-alternating Choquet capacities. These are set functions defined on the Borel algebra of a Polish (=complete, metrizable, separable) space, satisfying the following properties:

(a) $w(\varnothing)=0, \quad w(Y)=1$
(b) $w(A) \leqq w(B) \quad \forall A \subset B$
(c) $w(A_n) \uparrow w(A) \quad \forall A_n \uparrow A$
(d) $w(F_n) \downarrow w(F) \quad \forall \text{closed } F_n \downarrow F$
(e) $w(A \cap B) + w(A \cup B) \leqq w(A) + w(B)$

where all sets A, B, A_n are assumed Borel-measurable, and the whole set is denoted by Y. Huber and Strassen ([8], Sect. 2) show that there is a correspondence between 2-alternating capacities and certain *weakly compact* convex sets of probability measures – the condition (d) on w above being the equivalent of weak compactness of $\mathbf{Q}$.

The problem we wish to address is the assumption of compactness, which is necessary for the derivation of Huber-Strassen's robust Neyman-Pearson lemma. We would like to show in this note that compactness is indeed irrelevant for a large class of upper probabilities. The Huber-Strassen theorem in its present form does not even cover such simple cases as ε-contamination and the total variation norm since the resulting upper probabilities are capacities only on *compact* spaces Y [8, Ex. 3, 4], and the same holds for a class of upper probabilities which were introduced simultaneously and independently by Bednarski [1] and the present author [2]. As a consequence of our arguments in the following two sections, the restrictions to compact spaces in Huber-Strassen ([8], Ex. 3, 4) are not necessary, and we can get rid of *all* regularity conditions used by Bednarski ([1], Theorem 4.1) in his treatment of special capacities.

2. The Main Argument

It is easy to see that the upper probabilities derived from contamination and the total variation norm neighborhoods satisfy conditions (a), (b), (c), and (e), but not (d). However, we still have the following modified version (d′) of (d):

(d′) $w(A_n) \downarrow w(A) \quad \forall A_n \downarrow A \neq \varnothing$

where we only assume the sets A_n to be Borel-measurable as opposed to closed in (d). The reason (d′) holds is that the contamination or deviation in the total variation norm can sit in arbitrarily small non-empty sets, such that only the empty set escapes the possibility of hosting contamination or other deviating mass. For lack of a better name, we call a set function a *pseudo-capacity* if it satisfies conditions (a), (b), (c), (d′), and (e). For these pseudo-capacities, we can show the following:

Proposition. *On any Polish space Y, there exists a (possibly different) topology*
– which generates the same Borel algebra as the original Polish topology, and
– on which every pseudo-capacity is a proper capacity.

Proof. Below, we show that according to the Kuratowski isomorphism theorem, there exists a compact metrizable topology which generates the same Borel algebra as the original Polish topology on Y. On a compact space we have for any decreasing sequence of closed (hence compact) sets F_n: $F_n \downarrow \emptyset$ iff $F_n = \emptyset$ for some n. Hence, with $w(\emptyset)=0$ in mind, condition (d') results in (d) if specialized to closed sets.

There remains the proof of the existence of an equivalent metrizable compact topology. From the Kuratowski isomorphism theorem ([9], Theorem 2.12) it follows that two Polish spaces are measurably isomorphic iff they have the same cardinality, where an isomorphism is a bimeasurable bijection between the spaces. In any equivalence class of equicardinal Polish spaces, we can find an instance of a compact metric space:

- this is clear for finite sets;
- for countable sets, we can pick $\{1, 1/2, 1/3, \ldots\} \cup \{0\}$ with the absolute difference as a metric;
- and since uncountable Polish spaces have the cardinality of the continuum ([9], Theorem 2.8), the closed unit interval will do.

If we transfer such a compact metrizable topology to the space Y via a bimeasurable bijection, we obtain the required topology. □

3. The Application to Special Capacities

As mentioned in Sect. 1, Bednarski [1] and this author [2] independently defined a new type of set functions as follows: given

- a *concave continuous* function f: $[0, 1] \to [0, 1]$ satisfying $f(1)=1$, and
- a probability measure P on the Borel algebra of Y,

put $w(A)=f(P(A))$ for $A \neq \emptyset$, and $w(\emptyset)=0$. (One has to ask for continuity of f only at zero, everywhere else it follows from concavity, $f(1)=1$, and the range prescription $[0, 1]$. These conditions also force f to be non-decreasing.) Both Bednarski ([1], Lemma 3.1) and the author [2, 12.2] noticed:

Proposition. *The set function w is a pseudo-capacity; i.e., it satisfies conditions* (a), (b), (c), (d'), *and* (e).

Condition (a) is clear, (b) follows from monotonicity and (c) and (d') follow from continuity. Only condition (e) needs a calculation (which Bednarski [1], Sect. 3 omitted): for the numbers $u=P(A\cap B)$, $v=P(A\cup B)$, $x=P(A)$, and $y=P(B)$, we have $u+v=x+y$, $u \leqq x \leqq v$ and $u \leqq y \leqq v$. Hence we can write $x=\alpha u+(1-\alpha)v$ and $y=(1-\alpha)u+\alpha v$ for suitable $0 \leqq \alpha \leqq 1$. With concavity of f, this results in $f(u)+f(v) \leqq f(x)+f(y)$, and condition (e) follows. □

As mentioned in [1], Sect. 6, and [2], Sect. 12, we obtain contamination and total variation norm neighborhoods by specializing $f(x)=((1-\varepsilon)\cdot x+\varepsilon+\delta) \wedge 1$, which is clearly concave, continuous, has range $[0, 1]$, and satisfies $f(1)=1$.

According to the propositions of this and the previous section, these concave functions of probability measures are 2-alternating capacities, possibly in a non-standard topology. Such a topology, however, is enough to allow us to apply the Huber-Strassen theorem directly. We thus achieve considerable simplifications over Bednarski's results in that we can dispense with his regularity conditions ([1], 4.1), technical lemmas ([1], 3.3, 3.4), and his proof of the Huber-Strassen theorem ([1], 4.1) for these pseudo-capacities.

Another conclusion is that the author's results on robust finite decision problems and discrimination in [4], Sect. 5, 6 now hold on arbitrary Polish spaces. It was actually an oversight that the problem with the compactness requirement is not mentioned there, since the theory of that paper relies on the author's generalizations of the Huber-Strassen theorem to experiments with arbitrary finite parameter sets [3, Sect. 2, 8].

Finally, we would like to mention a curious aspect of any approach to pseudo-capacities based on the Huber-Strassen theorem: pseudo-capacities lead a life independent of topologies, as no topological concept enters their definition. It should therefore be possible to prove results of the Huber-Strassen type without resorting to topological arguments, very much in the spirit of Huber's earlier treatment of contamination and total variation norm neighborhoods [6, 7]. The fact that we may change the topology, if we like, is a strong hint that something of this sort can be done. On the other hand, a referee remarked that this is by no means clear: many measure theoretical properties (e.g., the existence of regular conditional probabilities) require topological assumptions, and the example of Huber's explicit solutions of some special problems make him rather sceptical in this regard. If he is right, then the results of this paper are the best we can hope for.

References

1. Bednarski, T.: On solutions of minimax test problems for special capacities. Z. Wahrscheinlichkeitstheorie Verw. Geb. **58**, 397–405 (1981)
2. Buja, A.: Sufficiency, least favorable experiments and robust tests. Unpublished PhD thesis, Swiss Federal Institute of Technology, ETH Zürich (1980)
3. Buja, A.: Simultaneously least favorable experiments, part I: upper standard functionals and sufficiency. Z. Wahrscheinlichkeitstheorie Verw. Geb. **65**, 367–384 (1984)
4. Buja, A.: Simultaneously least favorable experiments, part II: standard loss functions and their applications. Z. Wahrscheinlichkeitstheorie Verw. Geb. **69**, 387–420 (1985)
5. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Stat. **35**, 73–101 (1964)
6. Huber, P.J.: A robust version of the probability ratio test. Ann. Math. Stat. **36**, 1753–1758 (1965)
7. Huber, P.J.: Robust confidence limits. Z. Wahrscheinlichkeitstheorie Verw. Geb. **10**, 269–278 (1968)
8. Huber, P.J., Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. Ann. Stat. **1**, 251–263 (1973)
9. Parthasarathy, K.R.: Probability measures on metric spaces. New York-San Francisco-London: Academic Press 1967
10. Rieder, H.: Least favorable pairs for special capacities. Ann. Stat. **5**, 909–921 (1977)

Received September 12, 1985; in revised form April 16, 1986