KYBERNETIKA — VOLUME 31 (1995), NUMBER 3, PAGES 315–319

A REMARK ON EXISTENCE OF STATISTICAL FUNCTIONALS

IVAN MIZERA

An analytical condition for existence of an upper hemicontinuous extension of a statistical functional is given. It is demonstrated that it can be interpreted in statistical terms, namely, as a positive breakdown point property.

The use of statistical functionals, as a tool for establishing asymptotic properties of statistical procedures, can be traced back starting from the article by von Mises [10] — for references reflecting the further development see Reeds [11] and Fernholz [3]. However, the present author owes a credit for introducing him into the field to Hájek and Vorlíčková [4]. In Chapter 7 of this mimeographed lecture notes, intended to serve as an introductory statistics course, a concept of "naïve statistics" is introduced:

> "By a naïve statistics we shall call an approach to estimation of a parameter ϑ and its functions, based on the idea of replacing an unknown distribution law by a sample distribution law. ... If x_i are realizations of random quantities X_i, $1 \le i \le n$, the sample distribution is defined to be a discrete distribution which assigns to every point x_i the probability of $1/n$."[1]

Compared to the references quoted above, which are concerned with technical applications of statistical functionals, Hájek and Vorlíčková [4] emphasize a statistical aspect of the functional approach: the use of functionals for construction of statistical procedures.

In this note, we try to demonstrate how certain existence conditions for statistical functionals — although at first sight of purely analytic character — interact with statistical properties of functionals under consideration (or, perhaps better said, with statistical properties of underlying statistical procedures).

Particularly, the following problem is investigated: given a statistical functional $T: \mathcal{P}(X) \rightsquigarrow Y$ with $\mathrm{Dom}(T) \subset \mathcal{P}(X)$, when does its extension to a larger domain (ideally to whole of $\mathcal{P}(X)$) exist. By $\mathcal{P}(X)$, the set of all probabilities defined on the Borel sets of a Polish space X is denoted; Y is supposed to be a metric space.

[1] Translated from Czech by the author.

By $\mathrm{B}(A,\varepsilon)$, the set of all points with distance less than ε from A is denoted; if $A = \{x\}$, we write $\mathrm{B}(x,\varepsilon)$. A functional is, roughly, a function T from $\mathcal{P}(X)$ to Y — the slightly unusual notation $T\colon \mathcal{P}(X) \rightsquigarrow Y$ is adopted because T in our framework is allowed to be multi-valued; hence, formally, a *(statistical) functional* is a function defined on all of $\mathcal{P}(X)$ with values in 2^Y, the set of all subsets of Y. The brief look at some important fields of application — see Huber [7] for instance — provides a conviction that procedures which are either not defined at all, either not uniquely defined for certain arguments, constitute not just an exception. Multi-valued functions provide a convenient tool for covering these phenomena: the set of values can be empty or contain more than one element. Accordingly, the important sets connected with T are

$$\begin{aligned} \mathrm{Dom}(T) &= \{x \in X\colon \operatorname{card} T(x) > 0\}, \\ \mathrm{Uni}(T) &= \{x \in X\colon \operatorname{card} T(x) = 1\}. \end{aligned}$$

The image of a set $\mathcal{A} \subseteq \mathcal{P}(X)$ under T is defined to be

$$T(\mathcal{A}) = \bigcup_{P \in \mathcal{A}} T(P).$$

To preserve an informal character of the notation, for $x \in \mathrm{Uni}(T)$ we identify $T(x)$ with its single element, denoted also by $T(x)$.

Let $T\colon \mathcal{P}(X) \rightsquigarrow Y$ be a statistical functional. We say a functional $\tilde{T}\colon \mathcal{P}(X) \rightsquigarrow Y$ is an *extension* of T if $\mathrm{Dom}(T) \subset \mathrm{Dom}(\tilde{T})$ and $\tilde{T} = T$ on $\mathrm{Dom}(T)$. In the most common case, $\mathrm{Dom}(T) = \mathcal{E}(X)$, where $\mathcal{E}(X)$ is the set of all empirical probabilities; $P \in \mathcal{P}(X)$ is an empirical probability if it is a sample distribution in the sense of Hájek and Vorlíčková [4]: if $\partial(x)$ denotes the point ("Dirac") probability concentrated in x, then P can be written as

$$P = \partial(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \partial(x_i)$$

for some $x_1, x_2, \ldots, x_n \in X$. Usually, T arises from a sequence of estimators (our usage of this word embraces some other kinds of statistical procedures as well) $t_n\colon X^n \to Y$ which admit a functional representation

$$t_n(x_1, x_2, \ldots, x_n) = T(\partial(x_1, x_2, \ldots, x_n)).$$

A possible extension of T can be obtained using a concept of the asymptotic value — see Hampel [5]: $\tilde{T}(P)$ is set to be $t_\infty(P)$, where $t_\infty(P)$ is the set of points which are approached almost surely (or, alternatively, in probability) by any sequence $t_n(X_1, X_2, \ldots, X_n)$ such that $X_1, X_2, \ldots, X_n, \ldots$ are independent random elements, identically distributed according to the law P.

Although this approach is possible and in a sense minimal, in many cases it is more rewarding to use a (more) analytic approach: to seek an extension $\tilde{T}$ which is continuous with respect to a topology τ on $\mathcal{P}(X)$. Not only this approach fits to a rather well-studied realm, but it also yields a robustness property for underlying estimators — qualitative robustness; see Hampel [5] or Huber [7]. Moreover, if τ satisfies the law of large numbers — that is, for every sequence $X_1, X_2, \ldots, X_n, \ldots$ of independent random elements, identically distributed according to the law P, $\partial(x_1, x_2, \ldots, x_n)$ converges in τ to P almost surely — then $\tilde{T}(P) = t_\infty(P)$. Finally, the continuity approach can be used also in more general situations regarding the initial domain $\mathrm{Dom}(P)$. For weak topology, the law of large numbers is implied by the Varadarajan theorem (see Parthasarathy [12]); for various topologies similar to that generated by the Kolmogorov distance, by suitable extensions of the Glivenko-Cantelli theorem.

The framework of multi-valued functions calls also for an extended concept of continuity. A functional T is called *upper hemicontinuous* at $P \in \mathrm{Dom}(T)$ if for every $\varepsilon > 0$ there exists a neighbourhood U of P such that $T(U) \subseteq \mathrm{B}(T(P), \varepsilon)$. From the view of the current theory of multi-valued functions — see Aubin and Frankowska [1] — this is an "intermediate" definition: the stronger definition would use arbitrary open supersets of $T(P)$ instead of ε-neighbourhoods, the weaker one demands $T(P)$ to be only a set of limit points of $T(U)$ while U is shrinking to P. We deal overwhelmingly with the case when $T(P)$ is compact, in which all the definitions coincide. Note that if $T(P)$ is a singleton, upper hemicontinuity reduces to ordinary continuity: more precisely, $P_n \to P$ and $y_n \in T(P_n)$ imply $y_n \to T(P)$. Upper hemicontinuous functions are called also upper semicontinous; we prefer the prefix hemi- to avoid a possible confusion of $\subseteq$ order with $\leq$.

Every functional T possesses a trivial upper hemicontinuous extension: $\tilde{T}(P) = Y$ for every $P \in \mathcal{P}(X) \setminus \mathrm{Dom}(T)$. Hence, we are interested rather in minimal upper hemicontinuous extensions. A functional T is called *minimal upper hemicontinuous* at $P \in \mathrm{Dom}(T)$, if it is upper hemicontinuous and its value at P is minimal in this respect: there is no $A \subsetneq T(P)$ such that given any $\varepsilon > 0$, there exists a neighbourhood U of x with $T(U \setminus \{P\}) \subseteq \mathrm{B}(A, \varepsilon)$.

Proposition 1. Let $\mathcal{A}$ be dense in $\mathrm{Dom}(T)$. A functional $T\colon \mathcal{P}(X) \rightsquigarrow Y$, upper hemicontinuous at P, is minimal upper hemicontinuous at P if and only if for every $y \in T(P)$ there exists a sequence $P_n \in \mathcal{A}$ such that $P_n \to P$ and $T(P_n) \to y$.

Proof. Straightforward. □

Minimal upper hemicontinuous function is determined by its values on a set $\mathcal{A}$ dense in $\mathrm{Dom}(T)$. Quite often is $\mathrm{Uni}(T)$ dense in $\mathrm{Dom}(T)$. This slightly simplifies the picture.

Proposition 2. Suppose that $P \in \mathrm{Uni}(T)$. The following are equivalent:

- the restriction of T to $\mathrm{Uni}(T)$ is continuous at P (as an ordinary single-valued function);
- T is upper hemicontinuous at P;
- T is minimal upper hemicontinuous at P.

Proof. Straightforward. □

The necessary and sufficient criterion of Lechicki and Levi [9] — see also Lechicki and Levi [8], Holá [6] — employs the notion of subcontinuity. A functional is called *subcontinuous* at $P \in \mathcal{P}(X)$, if there exists an open set $U \subseteq \mathcal{P}(X)$ such that $P \in U$ and $T(U)$ is contained in a compact subset of Y. (Note that $P \in \mathrm{Dom}(T)$ is not required.)

Proposition 3. Let $T: \mathcal{P}(X) \rightsquigarrow Y$ be a closed-valued, upper hemicontinuous statistical functional. An upper hemicontinuous extension $\tilde{T}: \mathcal{P}(X) \rightsquigarrow Y$ compact-valued and minimal upper hemicontinuous at $\mathrm{Dom}(\tilde{T}) \setminus \mathrm{Dom}(T)$ exists if and only if T is subcontinuous at each $P \in \mathrm{Dom}(\tilde{T}) \setminus \mathrm{Dom}(T)$.

Proof. Follows from the Theorem 1.8 of Lechicki and Levi [9]. □

Now, suppose that the topology τ on $\mathcal{P}(X)$ is generated by a metric — for instance, for the weak topology it can be the Prokhorov metric. If Y is compact, subcontinuity trivially holds at every point. In the case of noncompact Y, the *breakdown point* of T is usually defined to be

$$\varepsilon_T^*(P) = \inf\{\varepsilon > 0\colon \text{ there is no compact } K \subseteq Y \text{ such that } T(\mathrm{B}(P,\varepsilon)) \subseteq K\},$$

in the functional setting — compare Hampel [5].

Proposition 4. A functional T is subcontinuous at P if and only if $\varepsilon_T^*(P) > 0$.

Proof. Straightforward. □

The breakdown point is known to be a robustness property; here we can see that it also influences existence properties of statistical functionals. Thus, starting from a functional defined on empirical probabilities, we can expect that a agreeable extension exists if the functional exhibits positive breakdown point. However, the following should be noted. The usual finite sample breakdown point, as defined by Donoho and Huber [2], is evaluated for P empirical; among the empirical probabilities of the same size; and with respect to the total variation metric. Hence, to have positive breakdown point in the sense of the definition given above, one should elaborate a possible extension (of a possibly existing breakdown result): general P (that we want to extend on) should be considered; all empirical probabilities are allowed as contaminations; and, possibly, a different metric, say, one of those satisfying the law of large numbers, should be ivolved (the last point may be simply dictated by an unsatisfactory features of the total variation topology). Nevertheless, in many cases such an extension turns to be quite straightforward.

Finally, it should be admitted that in a vast majority of cases, a suitable extension can be found directly, using a surrounding statistical background. Nevertheless, a purely existential criterion may be useful in more complex situations, which are of special interest particularly nowadays, in connection with introducing robust methods into more structured problems.

ACKNOWLEDGEMENT

The author is deeply grateful to Professor Marc Hallin and Institut de Statistique de l'Université Libre de Bruxelles for the hospitality and to Fonds d'encouragement à la recherche de l'U.L.B. for the support.

(Received December 5, 1994.)

REFERENCES

[1] J.-P. Aubin and H. Frankowska: Set-Valued Analysis. Birkhäuser, Boston 1990.
[2] D. L. Donoho and P. J. Huber: The notion of breakdown point. In: Festschrift in Honor of Erich Lehmann (K. Doksum and J. L. Hodges, eds.), Wadsworth, Belmont, CA.
[3] L. T. Fernholz: Von Mises Calculus for Statistical Functionals. (Lecture Notes in Statistics 19.) Springer, New York 1983.
[4] J. Hájek and D. Vorlíčková: Mathematical Statistics. Mimeographed lecture notes. SPN, Prague 1977 (in Czech).
[5] F. R. Hampel: A general qualitative definition of robustness. Ann. Math. Statist. *42* (1971), 1887–1896.
[6] L. Holá: An extension theorem for continuous functions. Czechoslovak Math. J. *38* (1988), 398–403.
[7] P. J. Huber: Robust Statistics. Wiley, New York 1981.
[8] A. Lechicki and S. Levi: Extensions of semicontinuous multifunctions. Istituto Lombardo (Rend. Sc.) *A 122* (1988), 79–85.
[9] A. Lechicki and S. Levi: Extensions of semicontinuous multifunctions. Forum Mathematicum *2* (1990), 341–360.
[10] R. von Mises: On the asymptotic distribution of differentiable statistical functions. Ann. Math. Statist. *18* (1947), 309–348.
[11] J. A. Reeds: On the Definition of von Mises Functionals. Ph. D. Dissertation, Harvard University, Cambridge, MA 1976.
[12] K. R. Parthasarathy: Probability Measures on Metric Spaces. Academic Press, New York 1967.

RNDr. Ivan Mizera, CSc., Department of Probability and Statistics, Comenius University, Mlynská dolina, 842 15 Bratislava. Slovak Republic.