Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga
o te Ūpoko o te Ika a Māui

How do we differentiate sets and why do we need it in statistics.

Estate Khmaladze
Victoria University of Wellington

Abstract

.

If, say, for each $t \geq 0$ we are given a set, usually a bounded set, K_t in $\mathbb{R}^d$, then we are given a set-valued function. One can think of it as a tube of variable shape, which evolves along the t-axis.

Examples are plentiful. A tea pot along its vertical axis, or along any other axis for this matter, defines a set-valued function. In obvious notations, any set of the form $B \cap \{T < t\}$ is a set-valued function. We, however, rarely speak about set-valued analysis in probability theory (with some notable exceptions like [1], [2], [9]) and so far – never in statistics.

In Lecture 1 we start with general description of the problem and heuristic approach to the notion of "fold-up" derivatives of set-valued functions. Our first question will be this: if N_n is, say, Poisson point process in $\mathbb{R}^d$ with intensity n, and if $K_t \to K_0$ when $t \to 0$, then what will be the limit of $N_n(K_t \triangle K_0)$, when $n \to \infty$ and $t \to 0$ simultaneously? Will it be a random variable which "lives" on the boundary ∂K_0 of the set K_0? The answer is "no". We will come to the conclusion that the limit should "live" on the so-called "generalized normal bundle" of differential geometry. Here and below our reference to the general geometric objects will be [11].

In Lecture 2 we will introduce and explain the use of further geometric objects, like the "local Steiner formula" and support measures. Then we will define the fold-up derivatives and show their basic properties (see [4]). In particular, we hope to discuss the relationship of this derivative with the theory of generalized functions. Although the final formulation is still to be written properly, we hope to show that fold-up derivatives lead to finer classification of the generalized functions than the existing theory.

In Lecture 3 we will discuss our motivational example: the change-set problem of statistics. Its basic formulation is as follows. Suppose we are given an i.i.d. sequence of observations $(X_i, Y_i)_{i=1}^n$, where $X_i \in \mathbb{R}^d$ are "locations", and Y_i are corresponding "marks", measured at these locations. Suppose there is a set, K, which we call "change-set", such that the distribution of Y_i given $X_i \in K$ is different from its distribution when $X_i \notin K$. This K is the parameter of interest and we want to estimate it or to test hypothesis about it. We will discuss also some other formulations

of the change-set problem and the results there which are relevant to our topic (see [5], [6], [7], [10]. In Lecture 3 – no differentiation.

We know that contiguity theory is a statistical theory on how to detect "small changes" in the parameter, or how to distinguish between hypothesis and its "local" alternaitves. The theory is developed for the problems where the parameter of interest is a vector or a function – these are objects which can be nicely imbedded in a "good" linear spaces. Sets do not enjoy this property. True that one can identify a set with its indicator function, but as we would show earlier in Lecture 2, this will not help us much. In Lecture 4 we will discuss how the notion of fold-up derivatives helps to build more or less complete analogue of the contiguity theory when the parameter is a set (references here are [3] and [8]).

References

[1] Z. Artstein. Invariant measures of set-valued maps, *J. Mathemat. Analysis Appl.* **252** (2000), 696-709.

[2] F. Bernardin. Multivalued Stochastic Differential Equation: Convergence of Numerical Scheme, *Set-Valued Analysis* **11** (2003), 393-415.

[3] J.H.J. Einmahl and E.V.Khmaladze, Central limit theorems for local empirical processes near boundaries of sets, *Bernoulli J.* (2010), accepted,.

[4] E.V. Khmaladze, Differentiation of sets in measure, *J. Mathemat. Analysis Appl.*, **334** (2007), No.2, 1055-1072.

[5] E.V. Khmaladze, R.M.Mnatsakanov, N.Toronjadze, Local covering numbers and change-set problem, *Mathemat. Methods Statistics*, (2006) **15**, 289-308

[6] E.V. Khmaladze, R.M.Mnatsakanov, N.Toronjadze, The change-set problem for VC-classes", *Mathemat. Methods Statistics,* (2006) **15**, 224-231.

[7] E.V. Khmaladze, N.Toronjadze, On a.s. coverage property of Voronoi tessellation", *Advances in Applied Probability*, (2001) **33**, 756-764

[8] E.V. Khmaladze and W.Weil, Local empirical processes near boundaries of convex bodies, *Annals Institute Statistical Mathematics* (2008), **60**, 813-842.

[9] B. Kyun Kim, Jai Heui Kim, Stochastic Integrals of Set-valued Processes and Fuzzy Processes, *J. Mathemat. Analysis Appl.*, **236** (1999), 480-502.

[10] M. D. Penrose, Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli J.* (2007), **13**, 1124-1150

[11] R. Schneider, *Convex Bodies: the Brunn-Minkowski Theory*, Cambridge University Press, Cambridge, 1993.

EVK