Differentiation of sets in measure

Estáte V. Khmaladze
Victoria University of Wellington
(School of Mathematics, Statistics and Computer Science, P.O.Box 600, Wellington, New Zealand, e-mail: Estate.Khmaladze@vuw.ac.nz)

Abstract. Suppose $F(\varepsilon)$, for each $\varepsilon \in [0,1]$, is bounded Borel subset of $\mathbb{R}^d$ and $F(\varepsilon) \to F(0)$ as $\varepsilon \to 0$. Let $A(\varepsilon) = F(\varepsilon) \triangle F(0)$ be symmetric difference and $\mathbb{P}$ be an absolutely continuous measure on $\mathbb{R}^d$. We introduce the notion of derivative of $F(\varepsilon)$ with respect to ε, $dF(\varepsilon)/d\varepsilon = dA(\varepsilon)/d\varepsilon$ such that

$$\frac{d}{d\varepsilon}\mathbb{P}(A(\varepsilon))|_{\varepsilon=0} = \mathbb{Q}(\frac{d}{d\varepsilon}A(\varepsilon)|_{\varepsilon=0}),$$

where $\mathbb{Q}$ is another, explicitly described, measure, although not in $\mathbb{R}^d$.

We discuss why this sort of derivative is needed to study local point processes in neighbourhood of a set: in short, if sequence of point processes $N_n, n = 1, 2, \ldots,$ is given on the class of set valued mappings $\mathcal{F} = \{F(\cdot)\}$ such that all $F(\varepsilon)$ converge to the same $F = F(0)$, then the weak limit of the local processes $\{N_n(A(\varepsilon)), F(\varepsilon) \in \mathcal{F}\}$ "lives" on the class of derivative sets $\{dF(\varepsilon)/d\varepsilon|_{\varepsilon=0}, F(\cdot) \in \mathcal{F}\}$.

We compare this notion of the derivative set-valued mapping with other existing notions.

Mathematical Subject Classification (1991). 28C99, 28E05, 46G05, 53C65, 60D05, 60G55.

Key words: local Steiner formula; local point process; set-valued mapping.

§1. Introduction

Consider a set-valued mapping $F(\varepsilon), \varepsilon \in [0,1]$ such that each $F(\varepsilon)$ is a bounded Borel subset of $\mathbb{R}^d$. Function-valued mappings, $f(\varepsilon, \cdot)$, being for each $\varepsilon \in [0,1]$ a function (from some measurable space X into, say, $\mathbb{R}$), are a very common object and we know, in particular, that the directional derivative in ε is again a function from X to $\mathbb{R}$. We would like to be able to say that the directional derivative of $F(\varepsilon)$ in ε is again a set, although not perhaps necessarily in $\mathbb{R}^d$. Denote $A(\varepsilon) = F(\varepsilon) \triangle F(0)$. Given a measure, $\mathbb{P}$, in $\mathbb{R}^d$ we would like also to give meaning to the formal equality

$$\frac{d}{d\varepsilon}\mathbb{P}(A(\varepsilon))|_{\varepsilon=0} = \mathbb{Q}(\frac{d}{d\varepsilon}A(\varepsilon)|_{\varepsilon=0}),$$

where $\mathbb{Q}$ is some other measure depending only on $\mathbb{P}$ and on the initial set $F(0)$ but not on the choice of mapping $F(\varepsilon)$.

The theory of set-valued mappings has not been used much in a probabilistic context so far [1]. It has been used in the statistical context even less – we know of no reference here. However, we will argue below that an extension of a well-known class of the statistical problems pertaining to the so called local empirical processes is essentially connected with the "local" analysis of set-valued mappings and naturally leads to the notion of its derivative.

The theory of set-valued mappings, rapidly developing in recent times, incorporates several notions of the derivative of $F(\varepsilon)$. Perhaps most general is the one when the derivative is understood as a family of tangent cones to the graph of the function $F(\varepsilon), \varepsilon \in [0,1]$, like contingent derivative of Aubin ([3]), Clarke derivative, and related notions. The corresponding theory is presented, e.g., in [3], Ch.4-5. Derivatives of a set-valued mapping when $F(\varepsilon)$ can be even a scalar function but ε takes values in, possibly, a complicated subset of $\mathbb{R}^d$ or in infinite-dimensional spaces, are given in [23], Ch.8; see also the fundamental survey paper [6].

In the papers [13] and [20] the notions of affine, semi-affine and ecliptic mappings are suggested in the role of differential mapping. In [26] the theory of quasi-affine mappings, as generalizations of affine and semi-affine mappings,

[1] We refer to the papers [2],[4],[19] as to important exceptions we know of.

was developed. The sets $F(\varepsilon)$ there are convex and bounded and, moreover, the graph of $F(\varepsilon), \varepsilon \in [0,1]$, is a convex set. A related but technically different notion of multi-affine mapping was introduced and studied in [1].

Another very beautiful approach of [22] and [28] suggests measures on the boundary ∂F, the Radon-Nikodym derivatives of $\mu_d(F(\varepsilon))$ with respect to $d\varepsilon$, as the derivatives of $F(\varepsilon)$. In [28] this approach is studied for the convex-valued mappings, i.e. when all $F(\varepsilon)$ are convex. It can be used in a more general set-up, as can be seen, in particular, in Theorem 1 below (see also short comments in §4 later).

The approach of the present paper is different from those mentioned above. It is based on the local Steiner formula, which connects Lebesgue measure of small "deformations" of a set F with so called support measures of the boundary ∂F. In this respect our basic reference is [24].

Our interest in differentiation of $F(\varepsilon)$ and the need to have a set as the derivative of $F(\varepsilon)$ stems from our attempt to develop the theory of local empirical processes in the neighbourhood of a given set, $F = F(0)$. The local empirical process in $\mathbb{R}^1$, that is, the empirical point process in the neighborhood of a point $c \in \mathbb{R}^1$ (or at ∞) appears in a very large number of statistical problems and forms a classical object of statistics. In multidimensional spaces the theory of local empirical process, again in the neighborhood of a point (or outside a large sphere) is a relatively recent development and we refer to the well-known papers [8], [9], [10], and, perhaps, [16], among others.

The local point process in the neighborhood of a set is a new object in statistical theory. To the best of our knowledge, the paper [18] is the first step in this direction. As far as a set is a more rich and diverse object than a point, the theory of local point processes in the neighborhood of a set promises to be more interesting and rich.

We describe these local processes and the need for set-differentiation in the separate short section.

In this paper we consider the case when $F(0)$ is convex body, whilst $F(\varepsilon)$ are more or less arbitrary. In particular, $F(\varepsilon)$ can be a union of disjoint components. Generalization to the case when $F(0)$ is the finite union of convex bodies is more or less clear – see, e.g. [24], Ch.4.4. The extension to the case when $F(0)$ can be any bounded set with "smooth" boundary (in technical terms – with a boundary of positive reach – see [12] or [24], p.212, or §2 below), or finite union of these, is immediate. The situation with arbitrary bounded F we hope to consider in later publications. This hope is connected with the recent results [14] on the existence of support measures and the local Steiner formula for arbitrary bounded F.

§2. Local Poisson processes

Consider a sequence $N_n, n = 1, 2, \dots,$ of Poisson point processes in $\mathbb{R}^d$ with intensity measure $EN_n(A) = n\mathbb{P}(A)$, where $\mathbb{P}$ is some given measure. As we know, for any Borel subset $A \subset \mathbb{R}^d$, $N_n(A)$ counts the number of random points in $\mathbb{R}^d$ that fall in A. Therefore as $n \to \infty$ the number of points in each given A of positive measure $\mathbb{P}$ grows unboundedly.
Let $\mathcal{V}_\varepsilon(\partial F)$ be a neighborhood of the boundary ∂F:

$$\mathcal{V}_\varepsilon(\partial F) = \{z \in \mathbb{R}^d : \|z - \partial F\| \le \varepsilon\}.$$

One can think of sets $A \subseteq \mathcal{V}_\varepsilon(\partial F)$ as describing "small" deviations from F: $A = F' \triangle F$ with F' depending on ε and tending to F as $\varepsilon \to 0$. Consider now a restriction of N_n to $\mathcal{V}_\varepsilon(\partial F)$,

$$N_{n\varepsilon} = \{N_n(A), A \subseteq \mathcal{V}_\varepsilon(F)\},$$

and let $n \to \infty$ and $\varepsilon \to 0$ simultaneously. This sequence of processes one can naturally call a *local Poisson process* in the neighborhood of the set F. The question is what is the limit of this process and where does this limit "live"? We will presently see that to think about the limiting process as living on the boundary ∂F is not satisfactory.

Suppose that $n\mathbb{P}(\mathcal{V}_\varepsilon(\partial F))$ converges to a constant as $n \to \infty$ and $n\varepsilon \to 1$. Then, for a given choice of $A(\varepsilon) \subseteq \mathcal{V}_\varepsilon(\partial F)$, it is natural to expect that $\varepsilon^{-1}\mathbb{P}(A(\varepsilon))$ converges to a finite limit, say, l, and therefore $N_n(A(\varepsilon))$ converges in distribution to a Poisson random variable N with intensity (expected value) l. However, it would be much nicer to be able to say that there is a set, say, B, not necessarily in $\mathbb{R}^d$ as we will argue below, and such that the limiting N is again a count of some other random points in this B, and that its intensity, being the limit l, is actually the value of some measure $\mathbb{Q}$ on B, $l = \mathbb{Q}(B)$. For statistical applications we would need to consider various classes of set-valued mappings $F(\varepsilon)$. If we could say that each $F(\varepsilon)$ has the derivative set B at $\varepsilon = 0$, then we will obtain the counting process $N(B)$ living on the class of the derivative sets, with intensity measure $\mathbb{Q}(B)$, and thus obtain the limiting object we would not have otherwise.

Existence of the limiting class of derivative sets is equally important for Gaussian limit theorems for the processes $N_{n\varepsilon}$, for laws of iterated logarithm and for other limit theorems.

To describe our notion of derivative sets in a simple situation consider the following examples of $A(\varepsilon)$. For the planar case, with $d = 2$, suppose that the boundary ∂F contains interval $\{a \le x \le b, y = 0\}$ with some $a < b$. Suppose the measure $\mathbb{P}$ is absolutely continuous and its density $p(x, y)$ is continuous in y at $y = 0$. Consider the sets

$$A(\varepsilon) = \{a \le x \le b, 0 \le y < \varepsilon g(x)\},$$

as $\varepsilon \to 0$, where g is, say, some positive bounded function. (We will consider broader class of functions in §5.) It is straightforward to derive the asymptotic

$$n\mathbb{P}\{A(\varepsilon)\} \sim \varepsilon^{-1} \int_a^b \int_0^{\varepsilon g(x)} p(x, y) dx dy \sim \int_a^b g(x) p(x, 0) dx dy,$$

and therefore to conclude that the $N_n(A(\varepsilon))$ converges in distribution to Poisson random variable with intensity equal to the right hand side above. However, if we consider the set

$$A^-(\varepsilon) = \{a \le x \le b, -\varepsilon g(x) \le y < 0\},$$

the asymptotics for $n\mathbb{P}\{A^-(\varepsilon)\}$ will be exactly the same. Since sets $A(\varepsilon)$ and $A^-(\varepsilon)$ are disjoint, the random variables $N_n(A(\varepsilon))$ and $N_n(A^-(\varepsilon))$ are independent and converge in distribution to independent Poisson random variables. Moreover, one can consider sets of the form $A_k(\varepsilon) = \{a \le x \le b, (k-1)\varepsilon g(x) \le y < k\varepsilon g(x)\}$ for different integer values of k and Poisson random variables $N_n(A_k(\varepsilon))$ for these values of k will all be independent and will converge to independent Poisson random variables with the same intensity. Therefore it will not be good to label all these limiting Poisson random variables by one and the same label – their intensity. What this paper says instead, however, is the following: there is a measure $\mathbb{Q}(dx, ds) = p(x, 0)dxds$, quite independent of the choice of $A_k(\varepsilon)$, and sets,

$$B_k = \{a \leq x \leq b, (k-1)g(x) \leq s < kg(x)\},$$

completely specified by these $A_k(\varepsilon)$, such that

$$\varepsilon^{-1}\mathbb{P}\{A_k(\varepsilon)\} \sim \mathbb{Q}(B_k).$$

These B_k we interpret as the derivatives of $A_k(\varepsilon)$ at $\varepsilon = 0$.

So far the derivative sets belonged to the same $\mathbb{R}^2$ as the sets $A_k(\varepsilon)$. However, as soon as instead of the interval $\{a \leq x \leq b, y = 0\}$, we have a (segment of) curve, it would not be really possible to stay in $\mathbb{R}^2$ and we would need to create an additional dimension. Indeed, suppose that F is a unit ball and suppose that x is a point on its boundary ∂F (which is a unit sphere). Let $A(\varepsilon) = \{z \in F : \|z - \partial F\| \leq \varepsilon T\}$. This is the set of the same nature as $A^-(\varepsilon)$ above but with the function g equal to constant T. Now if we "stretch" the $A(\varepsilon)$, as we did in the previous cases, for the values $T > 1$ there will be overlap, one-to-oneness will be lost, disjoint sub-sets will be mapped into overlapping subsets and the whole situation will become unnatural. Instead, the paper suggests to create "additional" dimension and consider the cylinder $\mathbb{R} \times \partial F$ and to map "stretched" subgraph on it. We will need to be slightly more careful about which cylinder to use, as it is explained in §3 and §4.

As we noted earlier, local Poisson process in the neighbourhood of a set was considered for the first time in [18] and for it the Poisson limit theorem, among other things, was proved without any use of the derivative sets (which the authors did not have at that time). This was possible because the process $N_{n\varepsilon}$ was considered on the whole σ-algebra of Borel subsets of $\mathcal{V}_\varepsilon(\partial F)$ and this σ-algebra was mapped on the Borel σ-algebra of the cylinder $\mathbb{R} \times \partial F$ (more precisely - of the cylinder Σ – see below). In this way asymptotic behavior of individual sequences $N_n(A(\varepsilon))$ and $A(\varepsilon)$ becomes not very important and not very visible. However, for the limit theorems on more restricted classes, like, e.g., Gaussian limit theorem, it is, in our view, unavoidable to develop the notion of derivative sets. In the last §6 we illustrate this type of limit theorems also in Poissonian case.

§3. Some preliminaries

Let F be a closed convex body in $\mathbb{R}^d$, that is, a closed convex set with interior points in $\mathbb{R}^d$. Let ∂F denote the boundary of F and let $P_{\partial F}(z)$ denote the metric projection of $z \in \mathbb{R}^d$ on ∂F, that is, $P_{\partial F}(z)$ is the nearest point to z from ∂F:

$$||z - P_{\partial F}(z)|| = \min_{x \in \partial F} ||z - x||.$$

The *skeleton* of ∂F is the set $S_{\partial F}$ defined as

$$S_{\partial F} = \{z \in \mathbb{R}^d : P_{\partial F}(z) \text{ is not unique}\}.$$

It is known that $\mu_d(S_{\partial F}) = 0$, where μ_d is Lebesgue measure in $\mathbb{R}^d$ (see [14]).

Let now $B^d(z, r)$ denote the closed ball with centre z and radius r. We will need to use the so called *local (interior) reach* $r(x), x \in F$ ([12]):

$$r(x) = \max\{r : B^d(z, r) \in F,\ B^d(z, r) \cap \partial F \ni x\}.$$

If $r(x) > 0$, the outer normal u to F at $x \in \partial F$ (with the norm $\|u\| = 1$) is unique, and $-u$ is its (unique) inner normal. Denote $Reg(F)$ the set of all points of ∂F which have unique outer normal. In general, however, at each $x \in \partial F$ there is a bundle of unit length outer normals which we denote $N(x)$.

The generalized *normal bundle* of F is defined as follows:

$$Nor(F) = \{(x, u) : x \in \partial F,\ u \in N(x)\}.$$

We use it to construct the cylinder $\Sigma = \mathbb{R} \times Nor(F)$ along with its subsets

$$\Sigma_+ = (0, \infty) \times Nor(F) \ \text{ and } \ \Sigma_- = (-\infty, 0] \times Nor(F).$$

For visualization purposes it will often be easier to consider the cylinder $\Gamma = \mathbb{R} \times Reg(F)$ and to project sets of Σ onto sets from Γ by letting $(t, x, u) \mapsto (t, x)$. For F with all its boundary points regular, $Reg(F) = \partial F$, one could use Γ from the very beginning. However, for general F this will be unsatisfactory as we need to control the contributions of "small" deformations of F in vicinity of irregular points of its boundary (see §3 below).

We define now the *local magnification map* τ_ε. Denote $d(z)$ the signed distance function

$$d(z) = \begin{cases} \|z - P_{\partial F}(z)\|, & \text{if } z \in \mathbb{R}^d \setminus F \\ -\|z - P_{\partial F}(z)\|, & \text{if } z \in F. \end{cases}$$

Then any point $z \in \mathbb{R}^d \setminus S_{\partial F}$ can be represented as

$$z = P_{\partial F}(z) + d(z)u,$$

with $u = (d(z))^{-1}(z - P_{\partial F}(z))$ if $d(z) \neq 0$. Now define

$$\tau_\varepsilon(z) = (\frac{d(z)}{\varepsilon}, P_{\partial F}(z), u), \;\; z \in \mathbb{R}^d \setminus S(\partial F)$$

Lemma 1. *(i) $\tau_\varepsilon(\cdot)$ maps the set $\mathbb{R}^d \setminus F$ onto Σ_+ :*

$$\tau_\varepsilon(\mathbb{R}^d \setminus F) = \Sigma_+$$

(ii) $\tau_\varepsilon(\cdot)$ maps the set $F \setminus S_{\partial F}$ into Σ_-:

$$\tau_\varepsilon(F \setminus S_{\partial F}) = \{(-\frac{r(x)}{\varepsilon}, 0] \times (x, u) : (x, u) \in Nor(F)\} \subseteq \Sigma_-$$

(iii) If $S_{\partial F}$ is nowhere dense then $\tau_\varepsilon(\cdot)$ is piecewise continuous and hence Borel measurable.

§4. Definition of differentiability

Consider a set-valued mapping $F(\varepsilon), 0 \leq \varepsilon \leq 1$, such that $F(0) = F$. With $F(\varepsilon)$ one can naturally associate "increments" $A^+(\varepsilon) = F(\varepsilon) \setminus F$, $A^-(\varepsilon) = F \setminus F(\varepsilon)$ and $A(\varepsilon) = A^+(\varepsilon) \cup A^-(\varepsilon) = F(\varepsilon) \triangle F$. It is natural to expect that the differentiability of $F(\varepsilon)$ at F is equivalent to the differentiability of $A(\varepsilon)$ at (as we prefer to say) ∂F.
We will use notation $\tau_\varepsilon(F(\varepsilon) \triangle F)$ for the image of the symmetric difference $F(\varepsilon) \triangle F$ in the local magnification map:

$$\tau_\varepsilon(F(\varepsilon) \setminus F) = B^+(\varepsilon) \subseteq \Sigma_+, \;\; \tau_\varepsilon(F \setminus F(\varepsilon)) = B^-(\varepsilon) \subseteq \Sigma_- \tag{1}$$

and the other way around

$$F(\varepsilon) = F \cup \tau_\varepsilon^{-1}(B^+(\varepsilon)) \setminus \tau_\varepsilon^{-1}(B^-(\varepsilon)).$$

The relationship between the subsets $F(\varepsilon)$ in $\mathbb{R}^d$ and their images in Σ established by (1) is shown in the following lemma.

Lemma 2. *Let F_1, F_2, F' be in $\mathbb{R}^d$ and B_1, B_2, B' be the images of the corresponding symmetric differences in Σ. If $B' = B_1 \cup B_2$, then*

$$F' = F \cup \tau_\varepsilon^{-1}(B_1^+ \cup B_2^+) \setminus \tau_\varepsilon^{-1}(B_1^- \cup B_2^-),$$

and if $B' = B_1 \cap B_2$, then

$$F' = F \cup \tau_\varepsilon^{-1}(B_1^+ \cap B_2^+) \setminus \tau_\varepsilon^{-1}(B_1^- \cap B_2^-).$$

The other way around: if $F' = F_1 \cup F_2$ then

$$B' = (B_1^+ \cup B_2^+) \cup (B_1^- \cap B_2^-),$$

and if $F' = F_1 \cap F_2$ then

$$B' = (B_1^+ \cap B_2^+) \cup (B_1^- \cup B_2^-).$$

Heuristically speaking this means that the union of sets in Σ "squeezes F' out of F", while the intersection of sets in Σ "pulls F' into F".

According to the *local Steiner formula* for any function f integrable with respect to the Lebesgue measure μ_d in $\mathbb{R}^d$

$$\int_{\mathbb{R}^d} f(z)\mu_d(dz) = \sum_{j=1}^{d} \binom{d-1}{j-1} \int_{Nor(F)} \int_{-r(x)}^{\infty} f(x+tu)t^{j-1}dt\theta_{d-j}(d(x,u)). \tag{2}$$

Here $\theta_{d-1}(A), \ldots, \theta_0(A)$ are finite Borel measures on $Nor(F)$ called *support measures* of F (see [24] for the theory of support measures). In particular, $\theta_{d-1}(\cdot)$ is Hausdorff measure on ∂F. For $f(z) = \mathbb{I}_A(z)$ being indicator function of a set $A \in \mathbb{R}^d$ we obtain

$$\mu_d(A) = \sum_{j=1}^{d} \binom{d-1}{j-1} \int_{Nor(F)} \int_{-r(x)}^{\infty} \mathbb{I}_A(x+tu)t^{j-1}dt\theta_{d-j}(d(x,u)). \tag{3}$$

Denote F_ε and $F_{-\varepsilon}$ the outer and inner parallel sets to the set F:

$$F_\varepsilon = F + \varepsilon B^d(0,1) \text{ and } F_{-\varepsilon} = F \overset{\star}{-} \varepsilon B^d(0,1)$$

(that is $F_{-\varepsilon} = \{z \in \mathbb{R}^d : z + \varepsilon B^d(0,1) \in F\}$). Denote M the measure on Σ defined as the direct product

$$M(ds, d(x,u)) = ds \times \theta_{d-1}(d(x,u)).$$

Definition 1. Call the (Borel) set-valued mapping $A(\varepsilon), 0 \le \varepsilon \le 1$, *differentiable* at ∂F at $\varepsilon = 0$ if for $\varepsilon \to 0$
(B) there exists finite $T > 0$ such that $\varepsilon^{-1}\mu_d(A(\varepsilon) \cap (F_{T\varepsilon} \setminus F_{-T\varepsilon})^c) \to 0$ and
(D) there exists a Borel set $B \in \Sigma$ such that $M(\tau_\varepsilon(A(\varepsilon))\triangle B) \to 0$.
Call the set B the *derivative* of $A(\varepsilon)$ at ∂F.

Definition 2. Call the (Borel) set-valued mapping $F(\varepsilon), 0 \le \varepsilon \le 1$, *differentiable* at F at $\varepsilon = 0$ if $F(\varepsilon)\triangle F$ is differentiable at ∂F. The *derivative* of $F(\varepsilon)$ at F is then defined to be the same as the derivative of $F(\varepsilon)\triangle F$ at ∂F. In notations

$$\frac{d}{d\varepsilon}F(\varepsilon)|_{\varepsilon=0} = \frac{d}{d\varepsilon}A(\varepsilon)|_{\varepsilon=0} = B.$$

The connection between the two definitions is, of course, the same as between the statements that $f(\varepsilon, \cdot)$ is differentiable if and only if the increment $f(\varepsilon, \cdot) - f(0, \cdot)$ is differentiable and both have the same derivative.

Note that B is not unique, but can be changed on a set of M measure 0. It allows, therefore, some manipulation with points on the boundary, for example.

The next lemma shows some algebraic properties of the differentiation.

Lemma 3. *(i) If $A_1(\varepsilon)$ and $A_2(\varepsilon)$ are differentiable at ∂F and B_1 and B_2 are corresponding derivatives, then $A_1(\varepsilon) \cup A_2(\varepsilon)$, $A_1(\varepsilon) \setminus A_2(\varepsilon)$ and $A_1(\varepsilon) \cap A_2(\varepsilon)$ are also differentiable at ∂F and the derivatives are $B_1 \cup B_2$, $B_1 \setminus B_2$ and $B_1 \cap B_2$ respectively.*
(ii) If $F_1(\varepsilon)$ is differentiable at F and $A_2(\varepsilon)$ is differentiable at ∂F and B_1 and B_2 are corresponding derivatives, then $F_1(\varepsilon) \cup A_2(\varepsilon)$ is differentiable at F and the derivative is B with $B^+ = B_1^+ \cup B_2^+$ and $B^- = B_1^- \setminus B_2^-$. At the same time $F_1(\varepsilon) \setminus A_2(\varepsilon)$ is also differentiable at F and the derivative is B with $B^+ = B_1^+ \setminus B_2^+$ and $B^- = B_1^- \cup B_2^-$.
(iii) For $a \in \mathbb{R}$ and $B \subseteq \Sigma$ define $aB = \{(as, x, u) : (s, x, u) \in B\}$. Let $f(\varepsilon)$ be continuous increasing function differentiable at 0 and $f(0) = 0$. If $F(\varepsilon)$ is differentiable, then $F(f(\varepsilon))$ is also differentiable and the derivative is $f'(0)B$.

Proof. (i) If T_1, T_2 are corresponding constants from condition (B) then $T = \max(T_1, T_2)$ is a suitable constant for $A_1(\varepsilon) \cup A_2(\varepsilon)$ and $A_1(\varepsilon) \cap A_2(\varepsilon)$, while T_1 is a suitable constant for $A_1(\varepsilon) \setminus A_2(\varepsilon)$: for this choice of constants the condition (B) will be satisfied. Consider condition (D). It is well-known (see, e.g. , [25], §2) that $M(\cdot\triangle\cdot)$ defines a (pseudo)metric in the class of Borel subsets of Σ. Hence, the condition (D) states that $B_i(\varepsilon) = \tau_\varepsilon(A_i(\varepsilon))$ converge to B_i, $i = 1, 2$, in this metric. However, τ_ε preserves set-theoretic operations, and these operations are continuous in the metric $M(\cdot\triangle\cdot)$. Namely, consider

$$\tau_\varepsilon(A_1^+(\varepsilon) \cup A_2^+(\varepsilon)) = B_1^+(\varepsilon) \cup B_2^+(\varepsilon)$$

and

$$\tau_\varepsilon(A_1^-(\varepsilon) \cap A_2^-(\varepsilon)) = B_1^-(\varepsilon) \cap B_2^-(\varepsilon).$$

If B_1 and B_2 denote the corresponding derivatives, then elementary inequalities

$$|\mathbb{I}_{B_1(\varepsilon)\cup B_2(\varepsilon)} - \mathbb{I}_{B_1\cup B_2}| \leq |\mathbb{I}_{B_1(\varepsilon)} - \mathbb{I}_{B_1}| + |\mathbb{I}_{B_2(\varepsilon)} - \mathbb{I}_{B_2}|$$

and

$$|\mathbb{I}_{B_1(\varepsilon)\cap B_2(\varepsilon)} - \mathbb{I}_{B_1\cap B_2}| \leq |\mathbb{I}_{B_1(\varepsilon)} - \mathbb{I}_{B_1}| + |\mathbb{I}_{B_2(\varepsilon)} - \mathbb{I}_{B_2}|$$

integrated with respect to measure $M(dt, d(x, u))$ lead to the result. Considerations for the difference $A_1(\varepsilon) \setminus A_2(\varepsilon)$ are similar.
(ii) Consider $(F_1(\varepsilon) \cup A_2(\varepsilon)) \setminus F = (F_1(\varepsilon) \setminus F) \cup (A_2(\varepsilon) \setminus F)$. Then from the statement (i) it follows, that the difference on the left hand side has the derivative $B^+ = B_1^+ \cup B_2^+$. Similarly, the equality $F \setminus (F_1(\varepsilon) \cup A_2(\varepsilon)) = (F \setminus F_1(\varepsilon)) \setminus A_2(\varepsilon)$ and statement (i) imply that the derivative of the left hand side exists and is equal to $B_1^- \setminus B_2^-$. Also, for $F_1(\varepsilon) \setminus A_2(\varepsilon)$ the equality $F_1(\varepsilon) \setminus A_2(\varepsilon) \setminus F = F_1(\varepsilon) \setminus F \setminus A_2(\varepsilon)$ and (i) imply that the derivative of the left hand side is $B_1^+ \setminus B_2^+$ while the equality $F \setminus (F_1(\varepsilon) \setminus A_2(\varepsilon)) = (F \setminus F_1(\varepsilon)) \cup (F \cap A_2(\varepsilon))$ and (i) implies that the derivative of the left hand side is equal to $B_1^- \cup B_2^-$.
(iii) To prove this statement it is sufficient to note that

$$\tau_\varepsilon(A(f(\varepsilon))) = \frac{f(\varepsilon)}{\varepsilon}\tau_{f(\varepsilon)}(A(f(\varepsilon))) = \frac{f(\varepsilon)}{\varepsilon}B(f(\varepsilon)).$$

Since $f(\varepsilon) \to 0$ as $\varepsilon \to 0$ the set $B(f(\varepsilon))$ converges to B. $\square$

Suppose $\mathbb{P}$ is an absolutely continuous measure in $\mathbb{R}^d$ and denote its density by p. Suppose also that on bounded sets $\mathbb{P}$ is finite. We would like to require that the density $p(z)$ can be approximated in the neighborhood of ∂F by a function depending on $P_{\partial F}(z)$ only. However, it is possible that the approximating functions are different for z tending to $P_{\partial F}(z)$ from outside

F and from inside F (cf. §6 below). Hence our formal requirement is that there are two functions $\bar{p}_+$ and $\bar{p}_-$ on ∂F, such that

$$\frac{1}{\varepsilon}\int_{F_\varepsilon\setminus F}|p(z)-\bar{p}_+(P_{\partial F}(z))|\mu_d(dz)\to 0$$
$$\frac{1}{\varepsilon}\int_{F\setminus F_{-\varepsilon}}|p(z)-\bar{p}_-(P_{\partial F}(z))|\mu_d(dz)\to 0 \tag{4}$$

Now define a measure $\mathbb{Q}$ on Σ as follows:

$$\mathbb{Q}(ds,d(x,u)) = ds\times\bar{p}_+(x)\theta_{d-1}(d(x,u)) \text{ for } s\geq 0,$$
$$\mathbb{Q}(ds,d(x,u)) = ds\times\bar{p}_-(x)\theta_{d-1}(d(x,u)) \text{ for } s< 0. \tag{5}$$

Let θ^c_{d-j} denote the part of θ_{d-j} absolutely continuous with respect to θ_{d-1}. Recall that these measures "live" on the set $Reg(F)$.

Theorem 4. *Suppose measure $\mathbb{P}$ satisfies condition* (4) *and suppose $\bar{p}_+$ and $\bar{p}_-$ are integrable with respect to $\theta^c_{d-j}, j=1,\dots,d$. If*

$$\varepsilon^{-1}\mathbb{P}(A(\varepsilon)\cap(F_{T\varepsilon}\setminus F_{-T\varepsilon})^c)\to 0 \text{ as } \varepsilon\to 0,$$

and if $A(\varepsilon)$ is differentiable at ∂F (with derivative $B\in\Sigma$) then

$$\frac{d}{d\varepsilon}\mathbb{P}(A(\varepsilon))|_{\varepsilon=0} = \mathbb{Q}(\frac{d}{d\varepsilon}A(\varepsilon)|_{\varepsilon=0})(=\mathbb{Q}(B)).$$

Remark 1. It is interesting to note that, as it can be seen from the proof below, the higher order measures $t^{j-1}dt\times\theta_{d-j}(d(x,u)), j=2,\dots,d$, do not have to be finite on the derivative set B.

Corollary 5. *Under conditions of the theorem*

$$\frac{d}{d\varepsilon}\mathbb{P}(F(\varepsilon))|_{\varepsilon=0} = \mathbb{Q}(\frac{d}{d\varepsilon}A^+(\varepsilon)|_{\varepsilon=0}) - \mathbb{Q}(\frac{d}{d\varepsilon}A^-(\varepsilon)|_{\varepsilon=0}).$$

Proof. It consists of establishing asymptotics for $\varepsilon^{-1}\mathbb{P}(A(\varepsilon))$. From the condition of the theorem it follows, that in doing so we can assume that $A(\varepsilon)\in F_{T\varepsilon}\setminus F_{-T\varepsilon}$. We can also put $T=1$. Consider an "intermediate" measure $\bar{\mathbb{P}}$ on $F_{T\varepsilon}\setminus F_{-T\varepsilon}$ with the density $\bar{p}_+(P_{\partial F}(z))$ or $\bar{p}_-(P_{\partial F}(z))$ according to $z\in F_{T\varepsilon}\setminus F$ or $z\in F\setminus F_{-T\varepsilon}$. Condition (4) implies that $\varepsilon^{-1}[\mathbb{P}(A(\varepsilon))-\bar{\mathbb{P}}(A(\varepsilon))]\to 0$ uniformly in $A(\varepsilon)$. Therefore we can consider

now only $\varepsilon^{-1}\bar{\mathbb{P}}(A(\varepsilon))$.
1) First consider $\varepsilon^{-1}\bar{\mathbb{P}}(A^-(\varepsilon))$. From the local Steiner formula it follows that

$$\bar{\mathbb{P}}(A^-(\varepsilon)) = \int_{Nor(F)} \int_{-\min(r(x),\varepsilon)}^{0} \bar{p}_-(x)\mathbb{I}_{A^-(\varepsilon)}(x+tu)dt\,\theta_{d-1}(d(x,u))+$$

$$\sum_{j=2}^{d} \binom{d-1}{j-1} \int_{Nor(F)} \int_{-\min(r(x),\varepsilon)}^{0} \bar{p}_-(x)\mathbb{I}_{A^-(\varepsilon)}(x+tu)t^{j-1}dt\,\theta_{d-j}(d(x,u)).$$

The sum of the higher order terms here is negligibly small. Indeed, for each integral we have

$$j\int_{Nor(F)} \bar{p}_-(x)|\int_{-\min(r(x),\varepsilon)}^{0} \mathbb{I}_{A^-(\varepsilon)}(x+tu)t^{j-1}dt|\,\theta_{d-j}(d(x,u))$$

$$\leq \int_{Nor(F)} \bar{p}_-(x)(\min(r(x),\varepsilon))^j\theta_{d-j}(d(x,u)) \leq \varepsilon^j \int_{Nor(F)} \bar{p}_-(x)\theta^c_{d-j}(d(x,u))$$

and the integral on the right hand side is finite. Therefore the sum is $O(\varepsilon^2)$.
2) It remains to see what is the asymptotic expression of the first summand. We have

$$\varepsilon^{-1}\int_{Nor(F)} \int_{-\min(r(x),\varepsilon)}^{0} \bar{p}_-(x)\mathbb{I}_{A^-(\varepsilon)}(x+tu)dt\theta_{d-1}(d(x,u)) =$$

$$= \int_{Nor(F)} \int_{-\min(r(x)/\varepsilon,1)}^{0} \bar{p}_-(x)\mathbb{I}_{B^-(\varepsilon)}(t,x,u)M(dt,d(x,u)).$$

However, with $B^-(\varepsilon) = \tau_\varepsilon(A^-(\varepsilon))$, the differentiability implies that the function $|\mathbb{I}_{B^-(\varepsilon)}(t,x,u) - \mathbb{I}_{B^-}(t,x,u)|$ tends to 0 $M-$a.e. on Σ_- and the Lebesgue majorised convergence theorem implies that

$$\int_{Nor(F)} \int_{-1}^{0} \bar{p}_-(x)|\mathbb{I}_{B^-(\varepsilon)}(t,x,u) - \mathbb{I}_{B^-}(t,x,u)|M(dt,d(x,u)) \to 0$$

3) Consider now $\varepsilon^{-1}\bar{\mathbb{P}}(A^+(\varepsilon))$. Again, from the local Steiner formula it follows that

$$\bar{\mathbb{P}}(A^+(\varepsilon)) = \sum_{j=1}^{d} \binom{d-1}{j-1} \int_{Nor(F)} \int_{0}^{\varepsilon} \bar{p}_+(x)\mathbb{I}_{A^+(\varepsilon)}(x+tu)t^{j-1}dt\theta_{d-j}(d(x,u))$$

However, the sum of the higher order terms here is again negligibly small, this time - without additional assumption on $\bar{p}_+$. Indeed,

$$\varepsilon^{-1}\sum_{j=2}^{d}\binom{d-1}{j-1}\int_{Nor(F)}\int_0^{\varepsilon}\bar{p}_+(x)\mathbb{I}_{A^+(\varepsilon)}(x+tu)t^{j-1}dt\theta_{d-j}(d(x,u))$$

$$\leq \frac{1}{d}\sum_{j=2}^{d}\binom{d}{j}\varepsilon^{j-1}\int_{Nor(F)}\bar{p}_+(x)\theta_{d-j}(d(x,u)) = O(\varepsilon)$$

because again, each integral on the right hand side is finite: applying condition (4) and local Steiner formula to $F_\varepsilon \setminus F$ we obtain

$$\mathbb{P}(F_\varepsilon \setminus F) \sim \bar{\mathbb{P}}(F_\varepsilon \setminus F) = \frac{1}{d}\sum_{j=1}^{d}\binom{d}{j}\varepsilon^{j-1}\int_{Nor(F)}\bar{p}_+(x)\theta_{d-j}(d(x,u))$$

and therefore all integrals indeed must be finite.
4) Asymptotic for the first summand follows in the same way as in 2). $\square$
Let $(x,u) \in Nor(F)$. The *section* of a set A by the line $z = x+tu$ (for $t \in \mathbb{R}$) is the set

$$A_{(x,u)} = \{z \in A : P_{\partial F}(z) = x, z - x \in \mathbb{R}u\}.$$

Similarly, the set

$$\tau_\varepsilon(A_{(x,u)}) = \tau_\varepsilon(A)_{(x,u)} = B_{(x,u)}$$

is the section of $B \in \Sigma$ by the line $\mathbb{R} \times (x,u)$.

Definition 3. Call the section $A_{(x,u)}(\varepsilon)$ of $A(\varepsilon)$ *differentiable* at $(x,u) \in Nor(F)$ at $\varepsilon = 0$ if for $\varepsilon \to 0$
(B) there exists $T > 0$ such that $j = 2,\ldots,d$

$$\varepsilon^{-1}\int(\mathbb{I}_{(-r(x),-\min(r(x),T\varepsilon)]}(t)+\mathbb{I}_{[T\varepsilon,\infty)}(t))\mathbb{I}_{A_{(x,u)}(\varepsilon)}(x+tu)t^{j-1}dt \to 0,$$

and
(D) there exists $B_{(x,u)} \in \mathbb{R} \times (x,u)$ such that

$$\int_{-r(x)/\varepsilon}^{\infty}\mathbb{I}_{B_{(x,u)}(\varepsilon)\triangle B_{(x,u)}}(s)ds \to 0.$$

Recall that $r(x)$ is the local reach of F at x.
Equivalent form of (B) is, of course, that for $j = 2, \dots, d$

$$\varepsilon^{j-1}(\int_{-r(x)/\varepsilon}^{-\min(r(x)/\varepsilon,T)} \mathbb{I}_{B_{(x,u)}(\varepsilon)}(s)s^{j-1}ds + \int_T^\infty \mathbb{I}_{B_{(x,u)}(\varepsilon)}(s)s^{j-1}ds) \to 0.$$

Remark 2. The role of the boundedness condition (B) may look in this definition somewhat peculiar. Indeed, let for simplicity $A_{(x,u)}$ be a subset of the ray $x + tu, t \geq 0$. For given x and u there is no sign of any presence of the set F or any Steiner formula associated with it, and therefore it may look strange to require anything except the proper differentiability condition (D), saying that the set $\varepsilon^{-1}A_{(x,u)}(\varepsilon)$ should "stabilize" as $\varepsilon \to 0$. However, we will need this condition later on when we "assemble" sections $A_{(x,u)}(\varepsilon)$ in one set $A(\varepsilon)$ around ∂F. Situation when (B) is not satisfied is discussed in the following example.

Example 1. Let $A(\varepsilon)$ be a "uniformly narrow" strip around ∂F, $A(\varepsilon) = F_{\alpha'(\varepsilon)} \setminus F_{-\alpha(\varepsilon)}$ with $\alpha'(\varepsilon), \alpha(\varepsilon) > 0$. Then each of its sections is the interval

$$A_{(x,u)}(\varepsilon) = \{z = x + tu : -\min(r(x), \alpha(\varepsilon)) < t \leq \alpha'(\varepsilon)\}$$

while

$$\tau_\varepsilon(A_{(x,u)}(\varepsilon)) = (-\varepsilon^{-1}\min(r(x), \alpha(\varepsilon)), \varepsilon^{-1}\alpha'(\varepsilon)] \times (x, u)$$

Therefore $A_{(x,u)}(\varepsilon)$ is differentiable if and only if $\varepsilon^{-1}\alpha(\varepsilon) \to a$ and $\varepsilon^{-1}\alpha'(\varepsilon) \to a'$, and the derivative set is $B_{(x,u)} = (-a, a'] \times (x, u)$. The mapping $A(\varepsilon)$ itself is differentiable under the same condition and the derivative is the set $B = (-a, a'] \times Nor(F)$. Both statements can be proved formally by isolating the set of (x, u) where the local reach $r(x)$ is small enough - just as was done in the proof of Theorem 4. The finite union of strips gives, in the present context, little new, but let us consider the countable union of strips, all outside F, say:

$$A(\varepsilon) = \cup_{i=1}^{\infty}(F_{a'_i\varepsilon} \setminus F_{a_i\varepsilon})$$

with $a_i < a'_i < a_{i+1}, i = 1, 2, \dots$. Then

$$A_{(x,u)}(\varepsilon) = \{z = x + tu : t \in \cup_{i=1}^{\infty}(\varepsilon a_i, \varepsilon a'_i]\}$$

and

$$\tau_\varepsilon(A_{(x,u)}(\varepsilon)) = \cup_{i=1}^{\infty}(a_i, a'_i] \times (x, u)$$

And this same set could be the derivative set, provided $\sum_{i=1}^{\infty}|a_i' - a_i| < \infty$. However, it may well be that this condition is satisfied and yet, for $j \geq 2$,

$$\frac{1}{\varepsilon}\int_{t\in\cup_{i=1}^{\infty}[\varepsilon a_i,\varepsilon a_i']} t^{j-1}dt = \varepsilon^{j-1}\frac{1}{j}\sum_{i=1}^{\infty}|(a_i')^j - a_i^j| = \infty$$

and (B) is violated. The condition (B) of Definition 1 is then also violated since

$$\varepsilon^{-1}\mu_d(A(\varepsilon)\cap F_{T\varepsilon}^c) =$$

$$= \theta_{d-1}(Nor(F))\sum_{i:a_i\geq T}|a_i' - a_i| + \frac{1}{d}\sum_{j=2}^{d}\binom{d}{j}\varepsilon^{j-1}\theta_{d-j}(Nor(F))\sum_{i:a_i\geq T}|(a_i')^j - a_i^j|$$

$$= \infty.$$

So, although there is the "stable" first term, the contribution from the higher order support measures is too high (infinite) and hence $A(\varepsilon)$ is not called here differentiable. This can not happen, however, if the sequence $\{a_i'\}_1^{\infty}$ is bounded, i.e. $A_{(x,u)}(\varepsilon) \subseteq F_{T\varepsilon}$.
The following theorem shows the connection between the differentiability of $A(\varepsilon)$ and of its sections in general. It is direct consequence of Fubini's theorem.

Theorem 6. *Suppose each integral in (B) of the Definition 3 is majorised by some function $\phi_j(x,u)$ integrable with respect to all measures θ_{d-j}. Then $A(\varepsilon)$ is differentiable (at ∂F at $\varepsilon = 0$) if sections $A_{(x,u)}(\varepsilon)$ are differentiable at θ_{d-1} -almost all $(x,u)\in\partial F$. The derivative of $A(\varepsilon)$ is the set*

$$B = \cup_{(x,u)\in Nor(F)}B_{(x,u)}$$

where $B_{(x,u)}$ is the derivative of $A_{(x,u)}(\varepsilon)$.

Proof. 1) For the inner integrals in (3) we obtain, from (B) of Definition 3,

$$\varepsilon^{-1}\int_{-r(x)}^{\infty}\mathbb{I}_{A_{(x,u)}(\varepsilon)}(x+tu)t^{j-1}dt \leq \varepsilon^{-1}\int_{-\min(r(x),T\varepsilon)}^{T\varepsilon}\mathbb{I}_{A_{(x,u)}(\varepsilon)}(x+tu)t^{j-1}dt + o(1)$$

$$\leq \frac{1}{j}\varepsilon^{j-1}2T^j + o(1) = o(1), \quad j\geq 2, \quad \varepsilon\to 0.$$

Since these integrals are also majorised by θ_{d-j}-integrable functions then

$$\frac{1}{\varepsilon}\sum_{j=2}^{d}\binom{d-1}{j-1}\int_{Nor(F)}\theta_{d-j}(d(x,u))\int_{-r(x)}^{\infty}\mathbb{I}_{A_{(x,u)}(\varepsilon)}(x+tu)t^{j-1}dt \to 0.$$

2) Since $(B(\varepsilon)\triangle B)_{(x,u)} = B_{(x,u)}(\varepsilon)\triangle B_{(x,u)}$ we have

$$\int \mathbb{I}_{B(\varepsilon)\triangle B}(t,x,u)dt\theta_{d-1}(d(x,u)) = \int \theta_{d-1}(d(x,u))\int \mathbb{I}_{B_{(x,u)}(\varepsilon)\triangle B_{(x,u)}}(t)dt$$

and the inner integral $\to 0$ because of the condition (D) of Definition 3. □

Example 2. Let Q_ε be some positive definite matrix, which tends to the identity matrix I as $\varepsilon \to 0$ and consider ellipsoids $F(\varepsilon) = \{x : x^tQ_\varepsilon x \le 1\}$. Then $F = F(0)$ is the unit ball. Obviously, the normal at $x \in \partial F$ is x itself. To find t such that $x + tu = (1+t)x \in \partial F(\varepsilon)$ we need to solve the equation $(1+t)^2x^tQ_\varepsilon x = 1$. Suppose that $Q_\varepsilon = I + \varepsilon D + o(\varepsilon)$. Then

$$t = -\frac{1}{2}\varepsilon x^tDx + o(\varepsilon)$$

and therefore the derivative of the section $A_{(x,x)}(\varepsilon)$, or rather projection of this derivative on Γ, is either $(0, -x^tDx/2] \times x$ or $(-x^tDx/2, 0] \times x$ depending on whether x^tDx is negative or positive. Theorem 6 implies that $F(\varepsilon)$ is differentiable and its derivative is the union of the sections above.

Example 3. Let again $F = F(0)$ be the unit ball, but this time assume that with ε increasing new "flanks" can branch away from it. Let again $d = 2$. We compare two cases, $F_i(\varepsilon) = F(0) \cup A_i(\varepsilon), i = 1, 2$. In the first case we set $F(\varepsilon) = \{x : (1-\varepsilon)x_1^2 + x_2^2 \le 1\}$ and choose $A_1(\varepsilon) = F(2\varepsilon) \setminus F(\varepsilon)$ as the strip between the two ellipsoids, while in the second one $A_2(\varepsilon) = B^d((1+\varepsilon)z_0, \varepsilon)$ is simply a shifted "small" ball. Here z_0 is a fixed unit vector.
Then, as it follows from Lemma 3, $F_1(\varepsilon)$ is differentiable. The mapping $F_2(\varepsilon)$ also is differentiable, but its derivative is set of measure 0. More precisely, the set $(\{0\} \times \partial F) \cup ((0, 2] \times z_0)$ is (the projection of) the limit of $\tau_\varepsilon(A_2(\varepsilon))$ in Hausdorff metric, but its M measure is 0. If we replace ε in $A_2(\varepsilon)$ by $\sqrt{\varepsilon}$ this will not improve the situation: the quotient $\varepsilon^{-1}\mu_d(A_2(\varepsilon))$ will have a finite limit, but there will be no limiting set for $\tau_\varepsilon(A_2(\varepsilon))$ in metric $M(\cdot\triangle\cdot)$.

Both $F_1(\varepsilon)$ and $F_2(\varepsilon)$ are differentiable at any other value of $\varepsilon > 0$.

§5. Further properties. Some discussion and examples.

"Deformations" $A(\varepsilon)$ as subgraphs.
One class of "small deformations" of the set F is naturally based on the notion of "small" functions, given on the normal bundle of F. Let $h_\varepsilon, \varepsilon \in [0,1]$, be a family of the functions on $Nor(F)$, which we will later assume small for small ε, and let h_ε^+ and h_ε^- be positive and negative parts of h_ε. Consider the sets in $\mathbb{R}^d$

$$\begin{aligned} A^+(h_\varepsilon) &= \{z \in \mathbb{R}^d \setminus F : 0 < d(z) \le h_\varepsilon^+(x,u)\} \\ A^-(h_\varepsilon) &= \{z \in F \setminus S_{\partial F} : -h_\varepsilon^-(x,u) < d(z) \le 0\} \\ A(h_\varepsilon) &= A^+(h_\varepsilon) \cup A^+(h_\varepsilon) \end{aligned} \tag{6}$$

where, as always, $x = P_{\partial F}(z)$ and u is the outer normal at x. One could call the set $A(h_\varepsilon)$ a subgraph of h_ε, but we rather reserve this term for its image $\tau_\varepsilon(A(h_\varepsilon))$. For a function g on $Nor(F)$, call the subsets of Σ defined as

$$g_{sub}^+ = \{(t,x,u) : 0 < t \le g^+(x,u)\}, \quad g_{sub}^- = \{(t,x,u) : -g^-(x,u) < t \le 0\}$$

and

$$g_{sub} = g_{sub}^+ \cup g_{sub}^-$$

the subgraphs of g^+, g^- and g respectively. Then

$$\begin{aligned} \tau_\varepsilon(A^+(h_\varepsilon)) &= g_{\varepsilon,sub}^+, \quad \text{where} \quad g_{\varepsilon,sub}^+ = \varepsilon^{-1} h_\varepsilon^+ \\ \tau_\varepsilon(A^-(h_\varepsilon)) &= g_{\varepsilon,sub}^-, \quad \text{where} \quad g_{\varepsilon,sub}^- = \varepsilon^{-1} \min(r(\cdot), h_\varepsilon^-). \end{aligned} \tag{7}$$

The next theorem connects differentiability of functions h_ε in ε with the differentiability of sets $A(h_\varepsilon)$.

Denote $\|h\|_j$ the norm of h in the space $\mathcal{L}_j(\theta_{d-j})$,

$$\|h\|_j = (\int_{Nor(F)} (h^+(x,u))^j \theta_{d-j}(d(x,u)))^{1/j} + (\int_{Nor(F)} (h^-(x,u))^j \theta_{d-j}^c(d(x,u)))^{1/j}$$

We say that h_ε is $\mathcal{L}_1$-differentiable if there is a function $g \in \mathcal{L}_1(\theta_{d-1})$ such that $\|\varepsilon^{-1} h_\varepsilon - g\|_1 \to 0$.

Theorem 7. *If* $\|h_\varepsilon\|_j = o(\varepsilon^{1/j}), j = 2, \ldots, d$, *then* $A(h_\varepsilon)$ *is differentiable if and only if* h_ε *is* $\mathcal{L}_1$*-differentiable. In this case*

$$\frac{d}{d\varepsilon} A(h_\varepsilon)|_{\varepsilon=0} = g_{sub}.$$

Remark 3. Similarly to Remark 1, we note that the conditions of the theorem allow the norms $\|g_\varepsilon\|_j$ to increase unboundedly although not too quickly: $\|g_\varepsilon\|_j = 0(\varepsilon^{-1+1/j})$. Consequently the limiting function does not have to have higher order norms $\|g\|_j, j = 2, \ldots, d$, finite. Actually, any function from $\mathcal{L}_1(\theta_{d-1})$ can be the limiting function.

Before we prove the theorem it seems convenient to single out the following statement as a separate lemma.

Lemma 8. *For* $g_1, g_2 \in \mathcal{L}_1(\theta_{d-1})$

$$M(g_{1,sub} \triangle g_{2,sub}) = \|g_1 - g_2\|_1.$$

Proof of Theorem 7. 1) According to (3)

$$\varepsilon^{-1}\mu_d(A(h_\varepsilon)) = \varepsilon^{-1}\int_{Nor(F)} (h^+_\varepsilon(x,u) + \min(r(x), h^-_\varepsilon(x,u))\theta_{d-1}(d(x,u)) + R(\varepsilon), \tag{8}$$

where the reminder term satisfies the inequality

$$\begin{aligned} R(\varepsilon) \leq & \frac{1}{d}\sum_{j=2}^{d}\binom{d}{j}\varepsilon^{-1}(\int_{Nor(F)} (h^+_\varepsilon(x,u))^j \theta_{d-j}(d(x,u)) \\ & + \int_{Nor(F)} (h^-_\varepsilon(x,u))^j \theta^c_{d-j}(d(x,u))) = o(1), \text{ as } \varepsilon \to 0 \end{aligned}$$

2) The integral in (8) above can be written as $M(g_{\varepsilon,sub})$ where g_ε is the function with positive and negative parts defined in (7). Now, if the limiting function g for $\varepsilon^{-1}h_\varepsilon$ exists then

$$|M(g_{\varepsilon,sub}) - M(g_{sub})| \leq M(g_{\varepsilon,sub} \triangle g_{sub})$$

and according to Lemma 8 (and triangle inequality)

$$M(g_{\varepsilon,sub} \triangle g_{sub}) = \|g_\varepsilon - g\|_1 \leq \|\varepsilon^{-1}h_\varepsilon - g\|_1 + \|\varepsilon^{-1}h_\varepsilon - g_\varepsilon\|_1.$$

The first norm on the right hand side tends to 0 by the condition and one can show (see the proof of Theorem 4 that the second norm also tends to 0. This ends the "if" part.
3) To prove the "only if" part we note that the differentiability of $A(h_\varepsilon)$ implies that the sets $g_{\varepsilon,sub}$ form Cauchy sequence in the metric $M(\cdot\triangle\cdot)$. Then, using Lemma 8, we see that the functions g_ε form Cauchy sequence: $\|g_{\varepsilon,sub} - g_{\varepsilon',sub}\|_1 \to 0$ as $\varepsilon, \varepsilon' \to 0$. But since the space $\mathcal{L}_1(\theta_{d-1})$ is complete, the limiting function $g \in \mathcal{L}_1(\theta_{d-1})$ exists. □

Shifts.
Let $F(\varepsilon) = F + \varepsilon A, A$ - a convex body. This mapping is called the affine mapping – see, e.g., [20]. Then $F(\varepsilon)$ is differentiable at F with the derivative

$$B^+ = s_A^+(\cdot)_{sub}, \quad B^- = s_A^-(\cdot)_{sub}$$

where s_A is the support function of A. A proof can be found, actually, in [24] , Ch1.7. Let, in particular, $F(\varepsilon) = F + \varepsilon a$ be a shift of the set F. Then again $F(\varepsilon)$ is differentiable at F with the derivative

$$B = \langle a, \cdot \rangle_{sub}.$$

More generally, one can formulate the following statement about the differentiability of the "smooth" shifts of differentiable mappings.

Definition 4. Call the section $A_{(x,u)}(\varepsilon)$ of $A(\varepsilon)$ *regularly differentiable* at $x \in Reg(\partial F)$ at $\varepsilon = 0$ if it is differentiable at $(x,u) \in Nor(F)$ and, for $z \in \partial F$, the following sets

$$\tilde{B}_{(z,u)}(\varepsilon) = \{s \in \mathbb{R} : z + \varepsilon su \in A(\varepsilon)\} \text{ and } B_{(x,u)}(\varepsilon) = \{s \in \mathbb{R} : x + \varepsilon su \in A(\varepsilon)\}$$

approximate each other:

$$\int_{-r(x)/\varepsilon}^{\infty} \mathbb{I}_{\tilde{B}_{(z,u)}(\varepsilon) \triangle B_{(x,u)}(\varepsilon)}(s) ds \to 0 \text{ as } z \to x, \, \varepsilon \to 0.$$

Call $A(\varepsilon)$ *regularly differentiable at* ∂F (and the corresponding $F(\varepsilon)$ *regularly differentiable at* F) if $A_{(x,u)}(\varepsilon)$ are regularly differentiable θ_{d-1} -a.e. on $Reg(\partial F)$.

Theorem 9. *Suppose $F(\varepsilon)$ is regularly differentiable at F with derivative set B and suppose the shift $a(\varepsilon) \in \mathbb{R}^d$ is such that $\varepsilon^{-1}a(\varepsilon) \to a' \in \mathbb{R}^d$. Then the mapping $F(\varepsilon) + a(\varepsilon)$ is differentiable at F and the derivative is $\bar{B}$ with*

$$\bar{B}^+ = (B^+ + \langle a', \cdot\rangle)^+ \cup \langle a', \cdot\rangle^+_{sub} \setminus (B^- + \langle a', \cdot\rangle)^+$$

and

$$\bar{B}^- = ((B^+ + \langle a', \cdot\rangle)^- \setminus \langle a', \cdot\rangle^-_{sub}) \cup (B^- + \langle a', \cdot\rangle)^-.$$

In the proof of this theorem we will need the following lemma.

Lemma 10. *Suppose $A(\varepsilon)$ is regularly differentiable at ∂F with the derivative B and suppose $a(\varepsilon) \in \mathbb{R}^d$ is such that $\varepsilon^{-1}a(\varepsilon) \to a' \in \mathbb{R}^d$. Then the mapping $A(\varepsilon) + a(\varepsilon)$ is differentiable at ∂F and the derivative is the set with the sections $B_{(x,u)} + \langle a', u\rangle$ for $x \in Reg(\partial F)$.*

Proof. As we know, the set of points of the boundary ∂F which are not regular have θ_{d-1} - measure 0. Therefore, in view of Theorem 6, it is sufficient, therefore, to define derivatives of sections $(A(\varepsilon)+a(\varepsilon))_{(x,u)}$ for $x \in Reg(\partial F)$. Suppose $z \in \partial F$ is such that $z + a(\varepsilon) = x + \lambda u$. For $x \in Reg(\partial F)$ and $\|a(\varepsilon)\| \sim \varepsilon\|a'\| \to 0$ this implies that $\lambda \sim \langle a(\varepsilon), u\rangle \sim \varepsilon\langle a', u\rangle$. Since the section we want is defined as

$$\{y : y = x + tu \in A(\varepsilon) + a(\varepsilon)\} = \{y : y - a(\varepsilon) = z + (t - \lambda)u \in A(\varepsilon)\},$$

we see that

$$(A(\varepsilon) + a(\varepsilon))_{(x,u)} = A_{(z,u)}(\varepsilon) + \langle a(\varepsilon), u\rangle$$

where, however, u is the normal at x but not necessarily the normal at z. Using regular differentiability condition we see that

$$\tau_\varepsilon((A(\varepsilon) + a(\varepsilon))_{(x,u)} = \tilde{B}_{(z,u)}(\varepsilon) + \varepsilon^{-1}\langle a(\varepsilon), u\rangle$$

can be approximated in measure by the set $B_{(x,u)}(\varepsilon) + \langle a', u\rangle$. □

Proof of Theorem 9. Since $F(\varepsilon) = A^+(\varepsilon) \cup F \setminus A^-(\varepsilon)$ we have $F(\varepsilon)+a(\varepsilon) = (A^+(\varepsilon) + a(\varepsilon)) \cup (F + a(\varepsilon)) \setminus (A^-(\varepsilon) + a(\varepsilon))$. However, both the $A^+(\varepsilon) + a(\varepsilon)$ and $A^-(\varepsilon) + a(\varepsilon)$ are differentiable at ∂F as it follows from Lemma 6, while $F + a(\varepsilon)$ is differentiable at F. Then Lemma 3 implies that $F(\varepsilon) + a(\varepsilon)$ is differentiable at F and the derivative is as stated in the theorem. □

Sets defined through inequalities. Quasi-affine mappings.
Suppose F is a polytope defined through the following minimal set of linear inequalities

$$F = \{x \in \mathbb{R}^d : \langle c_i, x\rangle \le b_i, i = 1, \dots, m\}$$

and let F_i be the $(d-1)$-dimensional face of F formed by the points $x \in \partial F$ such that $\langle c_i, x\rangle = b_i$. Let $F(\varepsilon)$ be also a polytope defined as

$$F(\varepsilon) = \{x \in \mathbb{R}^d : \langle c_i, x\rangle \le b_i + \varepsilon b_i', i = 1, \dots, m\}.$$

This $F(\varepsilon)$ is an affine mapping, and hence quasi-affine mapping and hence forms its own differential as it was defined in [26]. According to the definitions of the present paper this mapping is differentiable as well and the derivative is the set

$$B = g_{sub}, \quad \text{where} \quad g(x) = b_i', \ x \in F_i$$

Consider now general perturbation $F(\varepsilon)$ defined as

$$F(\varepsilon) = \{x \in \mathbb{R}^d : \langle c_i(\varepsilon), x\rangle \le b_i(\varepsilon), i = 1, \dots, m\}$$

where we only assume that vectors $c_i(\varepsilon)$ and scalars $b_i(\varepsilon)$ are differentiable at $\varepsilon = 0$: $c_i(\varepsilon) \sim c_i + \varepsilon c_i'$, $b_i(\varepsilon) \sim b_i + \varepsilon b_i'$. Although each $F(\varepsilon)$ is convex, the graph of it, $(F(\varepsilon), \varepsilon), \varepsilon \in [0, 1]$, does not have to be and typically is not convex in $\mathbb{R}^{d+1}$ even in the neighborhood of $(F(0), 0)$. Hence it is not quasi-affine. In Example 4 below we see that $F(\varepsilon)$ can not be approximated by a quasi-affine mapping with accuracy $o(\varepsilon)$ and therefore is not differentiable in the sense of [26]. However, the derivative of $F(\varepsilon)$ in the present meaning exists and can be described as follows (see the proof in [17]). Let $g_i(x) = b_i' - \langle c_i', x\rangle$ for $x \in F_i$ and $g_i(x) = 0$ for all other $x \in \partial F$. Then

$$\frac{d}{d\varepsilon} A(\varepsilon)|_{\varepsilon=0} = \cup_{i=1}^m g_{i,sub}.$$

Example 4. Let $d = 2$ and consider $F(\varepsilon) = \{0 \le x_1 \le 1, 0 \le x_2 \le 1 + \varepsilon x_1\}$. Certainly, each $F(\varepsilon)$ is convex, but the graph $(F(\varepsilon), \varepsilon), \varepsilon \in [0, 1]$, is not convex and neither can it be approximated with accuracy $o(\varepsilon)$ by a convex set. However, the derivative according to Definitions 1 and 2 of $F(\varepsilon)$ is g_{sub} where $g(x_1, x_2) = x_1$ for $x_2 = 1$ and $g(x_1, x_2) = 0$ for all other points of the boundary of $F(0)$.

As to the inverse question whether quasi-affine mappings are differentiable in the present meaning, the answer is positive as the following proposition shows. One of the key points of the proof was actually proved in [26].

Let a quasi-affine mapping be defined as

$$F(\varepsilon) = \cap_{\|\psi\|=1}\{x \in \mathbb{R}^d : \langle x, \psi\rangle \leq s(\psi) + \varepsilon c(\psi)\}$$

where $s(\cdot)$ is support function of the set F and $c(\cdot)$ is some positively homogeneous function. For a given quasi-affine mapping this function is not unique and can be very different from $\varepsilon^{-1}(s(\varepsilon,\cdot) - s(\cdot))$, where $s(\varepsilon,\cdot)$ denotes the support function of the set $F(\varepsilon)$ (cf. [26], Sec.2 and Lemma 2.14 in particular).

Theorem 11. *A quasi-affine mapping is differentiable in the sense of Definition 2.*

Corollary. *A set-valued mapping* $F(\varepsilon), \varepsilon[0,1]$, *differentiable in the sense of Definition 3.1 of [26], i.e. such that there exists a quasi-affine mapping* $F'(\varepsilon), \varepsilon[0,1]$, *which approximates* $F(\varepsilon)$ *in Hausdorff metric with the rate* $o(\varepsilon)$, *is differentiable in the sense of Definition 2.*

Proof. We will construct sections of the derivative set at any regular point of the boundary of F and then use Theorem 6. Denote $c_*(\varepsilon,\cdot) = \varepsilon^{-1}(s(\varepsilon,\cdot) - s(\cdot))$. For $x \in Reg(\partial F)$ let, as usual, u denote its (unique) outer normal, and let λ be such that $x + \lambda u \in \partial F(\varepsilon)$. Since u is normal at x and hence $\langle x, u\rangle = s(u)$, the inequality $\langle x+\lambda u, u\rangle \leq s(\varepsilon, u)$ leads to $c_*(\varepsilon, u) \geq \varepsilon^{-1}\lambda$. At the same time, there is a supporting hyperplane through $x + \lambda u$ and hence $\psi = \psi_\varepsilon$ such that $\langle x + \lambda u, \psi\rangle = s(\varepsilon, \psi)$ and hence

$$\varepsilon^{-1}\lambda = \frac{s(\psi) - \langle x, \psi\rangle}{\varepsilon\langle u, \psi\rangle} + \frac{c_*(\varepsilon, \psi)}{\langle u, \psi\rangle}$$

Since $\langle x, \psi\rangle \leq s(\psi)$ the latter equality leads to inequality

$$\varepsilon^{-1}\lambda \geq \frac{c_*(\varepsilon, \psi)}{\langle u, \psi\rangle}.$$

Now, as $\varepsilon \to 0$, $c_*(\varepsilon,\cdot)$ forms a non-decreasing (in ε) sequence of continuous functions in ψ, bounded from above by $c(\cdot)$ (see Lemma 2.14 of [26]) and hence it converges to some function $c_*(\cdot)$ uniformly in ψ:

$$\sup_{\|\psi\|=1} |c_*(\varepsilon, \psi) - c_*(\psi)| \to 0.$$

However, since $\psi \to u$ as $\varepsilon \to 0$ we see that $\varepsilon^{-1}\lambda \to c_*(u)$. Now note that the interval $(0, \varepsilon^{-1}\lambda]$ (if $\varepsilon^{-1}\lambda > 0$ and the interval $(\varepsilon^{-1}\lambda, 0]$ if $\varepsilon^{-1}\lambda < 0$) is the section $B_{(x,u)}(\varepsilon)$, and we proved that these sections converge at any $x \in Reg(\partial F)$. The rest follows from Theorem 6. □

As we said in the introduction, the notion of multi-affine mapping, in the role of differential mapping, was introduced and studied in [1]. For one-dimensional ε, it is the mapping defined as

$$F(\varepsilon) = \cup_{b \in B}\{\varepsilon b + D_b\}$$

where D_b, for each b, and B are subsets of $\mathbb{R}^d$. According to Definition 4.1 of [1], the mapping $F'(\varepsilon)$ is directive (differentiable) if there exists a multi-affine mapping $F(\varepsilon)$ which approximates $F'(\varepsilon)$ in Hausdorff metric with the rate $o(\varepsilon)$ - the property, as Z.Artstein points out in §9 of [1], useful in various applications of the notion. We do not go into study of this very attractive notion in any significant detail here, but merely note that the similar property can be noted about the mappings differentiable in the sense of the present paper. Namely, let $\tau_\varepsilon^{-1}(B) = \{z \in \mathbb{R}^d : \tau_\varepsilon(z) \in B\}$. Then, if $A(\varepsilon)$ is differentiable at ∂F and B is the derivative set, then

$$|\mu_d(A(\varepsilon)) - \mu_d(\tau_\varepsilon^{-1}(B)| = o(\varepsilon).$$

Derivatives as measures.
Let us make a brief comment on how the approach of [28] relates to the present one. Condition (D) of Definition 2 implies, that $\mu_d(A(\varepsilon))$ is absolutely continuous in ε (at $\varepsilon = 0$), which is essential point used in [22] and [28] as well. In particular, the function $\mu(B_{(x,u)})$ is the density function used in [28], p.340. In the context of our problem with the local point processes, one could, in principle, agree to label the limiting process by these density functions, if not the following consideration: there are many different and unrelated sequences of shrinking sets $A(\varepsilon)$ which would lead to the same density function. As we said in §2, let $N_n(A), A \in F_\varepsilon \setminus F_{-\varepsilon,}$ be a "local" Poisson point process, and suppose its intensity measure is $n\mu_d(A)$, and let $n \sim \varepsilon^{-1}$. Take $A_1(\varepsilon) = \tau_\varepsilon^{-1} g_{sub}$ and, to avoid technicalities associated with

the local reach, assume that $g(x,u) > 0$. Let $A_2(\varepsilon) = \tau_\varepsilon^{-1}(2g_{sub} \setminus g_{sub})$. Then the limiting density functions for both cases will be the same and equal to $g(x,u)$. However, $A_1(\varepsilon)$ and $A_2(\varepsilon)$ are disjoint sets and with each of them Poisson random variables, $N_n(A_1(\varepsilon))$ and $N_n(A_2(\varepsilon))$, are associated and these two random variables are independent. Moreover, one can construct as many such independent random variables as one wishes by considering $A_m(\varepsilon) = \tau_\varepsilon^{-1}(mg_{sub} \setminus (m-1)g_{sub})$ all with the same limiting density function $g(x,u)$. It will be unsatisfactory to "glue up" these random variables in the limit. Definitions 1 and 2 allows one to avoid this and to separate the set B and the measure M.

Derivatives as tangent cones. (Connections with contingent derivatives of J.-P. Aubin and Clarke's derivative.)
Definition of the derivative of a set-valued mapping through tangent cones to its graph is, as we said in the introduction, very general and well developed. It also is based on a lucid geometric idea. Namely, let $\{(\varepsilon, y) \in [0,1] \times \mathbb{R}^d : y \in F(\varepsilon)\}$ be the graph of $F(\varepsilon), \varepsilon \in [0,1]$ and let $x \in \partial F(0)$. Suppose $T_{(0,x)}$ is a tangent cone to this graph at the point $(0,x)$ (see Ch 5, [3]). Then the set-valued mapping $DF_{(0,x)}(\eta), \eta \in [0,1]$, defined as

$$DF_{(0,x)}(\eta) = \{y \in \mathbb{R}^d : (\eta, y) \in T_{(0,x)}\}$$

is called derivative mapping of $F(\varepsilon), \varepsilon \in [0,1]$.
To illustrate the connections between $DF_{(0,x)}(\cdot)$ and the derivative $dF(\varepsilon)/d\varepsilon|_{\varepsilon=0}$ suggested in this paper consider the following simple example. This example will also show the difference between the two notions.
Let $d = 2$ and let $F(\varepsilon) = \{x = (x_1, x_2) : x_1, x_2 \geq 0, \|x\| \leq 1 - \varepsilon\}$. The graph of this set-valued mapping has tangent cone at each boundary point $(\varepsilon, x), x \in \partial F(\varepsilon)$. Recall that this tangent cone is defined, say, for $\varepsilon = 0$ as

$$T_{(0,x)} = \{z \in \mathbb{R}^3 : \|hz - (0,x)\| = o(h), h \to 0\}.$$

It is not very important now to stress that Clarke tangent cone, with more stringent definition, also exists in this example at any boundary point. What is probably more important is to note that $T_{(0,x)}$ is not a tangent hyperplane alone. In particular, if x is regular point of $\partial F(0)$, i.e. if there is unique outer normal u, then for any $\eta \in [0,1)$

$$DF_{(0,x)}(\eta) = \{y \in \mathbb{R}^d : \langle y, u \rangle \leq 1 - \varepsilon\}$$

(and not only $\{y \in \mathbb{R}^d : \langle y, u\rangle = 1 - \varepsilon\}$).
The derivative set B for this mapping also exists and its sections $B_{(x,u)}$ at these x can be connected with $DF_{(0,x)}$ as follows:

$$B_{(x,u)} = B^-_{(x,u)} = \{h \in \mathbb{R} : \eta h u \in (F - x) \setminus DF_{(0,x)}(\eta)\}.$$

(For the mapping defined as $F(\varepsilon) = \{x = (x_1, x_2) : x_1, x_2 \geq 0, \|x\| \leq 1 + \varepsilon\}$ we would have

$$B_{(x,u)} = B^+_{(x,u)} = \{h \in \mathbb{R} : \eta h u \in DF_{(0,x)}(\eta) \setminus (F - x)\}.)$$

However, for non-regular points of $\partial F(0)$, with more than one outer normal, as, for example, for $x = (1, 0)$, the situation is different. The sets $DF_{(0,x)}(\eta)$ are still uniquely defined, while the sections of the derivative set B are not defined, or not defined uniquely.
These non-regular points of the boundary can actually be most interesting points in many optimization problems, and it is important to have a notion of the derivatives, like $DF_{(0,x)}(\cdot)$, equally applicable to regular and non-regular boundary points. However, the fact that the derivative sets B are formed basically by sections $B_{(x,u)}$ at regular points x comes not from attempt to simplify or trivialize the approach. It stems from another fact that linear changes of order ε in the neighborhood of all non-regular points of the boundary lead only to changes of order ε^2 or higher in the measure and therefore are indeed negligible in the asymptotics of the first order.

We suppose that sets of non-regular points, to which higher order support measures attach non-zero mass, will find a natural place as part of higher order derivatives, whatever these derivatives may prove to be. The reader may agree with this supposition observing that, for example, the second derivative of Lebesgue measure of the set $A^+(\varepsilon)$ as defined in (6) naturally would be

$$\frac{d^2}{d\varepsilon^2}\mu_d(A^+(\varepsilon))|_{\varepsilon=0} = \int \frac{d^2}{d\varepsilon^2} h^+_\varepsilon(x, u)|_{\varepsilon=0}\, \theta_{d-1}(d(x, u))$$

$$+ \int \frac{d}{d\varepsilon} h^+_\varepsilon(x, u)|_{\varepsilon=0}\, \theta_{d-2}(d(x, u))$$

and therefore incorporates the next support measure.

§6. Convergence of the local Poisson process

Whenever in a problem of statistical inference a set becomes the parameter of interest the local analysis with respect to this set will be needed. Indeed, we know that asymptotic statistical theory is very much based on the local behaviour of the likelihood process, both for the so called *parametric problems*, when the parameter of interest is a point in $\mathbb{R}^d$ (see, e.g., [15]), and for *semi-parametric* problems, when the parameter of interest is a function (see, e.g., [5]). It should be no less true when the parameter is a set.

As a particular example of such problems one can consider the class of so-called *change-set* problems of spatial statistics. (These problems are also connected with *image analysis*.) In a simple version of this problem, assume that within a certain region (an "image") $K \subset \mathbb{R}^d$ the intensity of Poisson process N_n is nc_1, while outside K it is nc_0 with a different constant c_0 (cf., e.g., [7], while the more general version with discontinuities in the so-called *regression function*, although for one-dimensional time, can be found in [21]). The region K is unknown and the inference about K should be made from the "observation" N_n.

Suppose we wish to test the basic (null) hypothesis that $K = F$ for some given $F = F(0)$, while under the alternative hypothesis K can be any member of some given class $\mathcal{F}(\varepsilon)$ of "deviations" $F(\varepsilon)$ from F. The basis for discrimination between the two hypotheses is provided by the so-called *log-likelihood*, which in this problem has the form

$$L_n = \ln \frac{c_1}{c_0}[N_n(F(\varepsilon)\setminus F) - N_n(F\setminus F(\varepsilon))] - n(c_1 - c_0)[\mu_d(F(\varepsilon)\setminus F) - \mu_d(F\setminus F(\varepsilon))]$$

For large n, discrimination between F and "quite" distinct $F(\varepsilon)$ can become easy and the theory should focus, as in parametric and semi-parametric cases, on the asymptotics of L_n when $F(\varepsilon) \to F$ along with $n \to \infty$. Thus we are led not to one but to a class of set-valued mappings all converging to the same F as $\varepsilon \to 0$. Then L_n becomes simply a version of the local Poisson process introduced in §2. This is, essentially, true for more intricate formulations of the change-set problem (cf., e.g., [18]).

Suppose $\mathcal{F}(\varepsilon) = \{F_\gamma(\varepsilon), \gamma \in \Gamma\}$ is the alternative class of change-sets and let

$\mathcal{A}(\varepsilon) = \{A_\gamma(\varepsilon) = F_\gamma(\varepsilon) \triangle F, \gamma \in \Gamma\}$. Slightly modifying the notation of §2 let

$$N_{n\varepsilon} = (N_n, \mathcal{A}(\varepsilon)) = \{N_n(A), A \in \mathcal{A}(\varepsilon)\} \tag{9}$$

and recall that $EN_n(A) = n\mathbb{P}(A)$. Now suppose each $A_\gamma(\varepsilon)$ is differentiable at ∂F and denote

$$\frac{d}{d\varepsilon}\mathcal{A}(0) = \{\frac{d}{d\varepsilon}A_\gamma(\varepsilon)|_{\varepsilon=0}, \gamma \in \Gamma\}$$

On this class of sets, in Σ, introduce now the Poisson process

$$N = (N, \frac{d}{d\varepsilon}\mathcal{A}(0)) = \{N(B), B \in \frac{d}{d\varepsilon}\mathcal{A}(0)\} \tag{10}$$

with intensity measure $EN(B) = \bar{c}\,\mathbb{Q}(B)$ with some constant $\bar{c}$. Our aim is to show that the current notion of differentiability naturally places N as the limiting process for $N_{n\varepsilon}$.

Theorem 12. *If $n \to \infty$ and $n\varepsilon \to \bar{c}$ then all finite dimensional distributions of the process* (9) *converge in total variation to the corresponding finite dimensional distributions of the process* (10). *In notation*

$$(N_n, \mathcal{A}(\varepsilon)) \xrightarrow{f.d.t.v.} (N, \frac{d}{d\varepsilon}\mathcal{A}(0))$$

Proof. Differentiability assumption on $A_\gamma(\varepsilon)$ implies that, for every finite m, we can assume that $A_{\gamma_j} \subseteq \mathcal{V}_{T\varepsilon}(\partial F)$ with one common $T = T(m)$. Denote, within this proof, $A^0(\varepsilon) = \mathcal{V}_{T\varepsilon}(\partial F) \setminus A(\varepsilon)$ while $A^1(\varepsilon) = A(\varepsilon)$. Denote $\Omega_m = \Omega$ collection of all vectors $\omega = (\omega_1, \dots, \omega_m)$ with each ω_j being 0 or 1 and consider pairwise disjoint sets

$$C_\omega(\varepsilon) = \cap_{j=1}^m A_{\gamma_j}^{\omega_j}(\varepsilon),\ \omega \in \Omega\,.$$

The distribution, in $\mathbb{R}^m$, of $\{N_n(A_{\gamma_j}(\varepsilon)), j = 1, \dots, m\}$ is uniquely determined by the distribution of $\{N_n(C_\omega(\varepsilon)), \omega \in \Omega\}$. Then the rest of the proof follows from the two facts: Lemma 3 implies that each $C_\omega(\varepsilon))$ is differentiable at ∂F with derivative $D_\omega = \cap_{j=1}^m B_{\gamma_j}^{\omega_j}$ where $B^0 = \Sigma_T \setminus B$ and $B^1 = B$, while Theorem 4 implies that

$$\sum_{\omega \in \Omega} |n\mathbb{P}(C_\omega(\varepsilon)) - \bar{c}\,\mathbb{Q}(D_\omega)| \to 0 \tag{11}$$

Indeed, suppose $\mathcal{P}_{n\varepsilon,m}$ and $\mathcal{P}_m$ are two Poisson distributions in $\mathbb{R}^m$ corresponding to $\{N_n(C_\omega(\varepsilon)), \omega \in \Omega\}$ and $\{N(D_\omega(\varepsilon)), \omega \in \Omega\}$ respectively. Then the distance in variation between $\mathcal{P}_{n\varepsilon,m}$ and $\mathcal{P}_m$ is

$$E|\frac{d\mathcal{P}_{n\varepsilon,m}}{d\mathcal{P}_m}(N) - 1| \tag{12}$$

where the Radon-Nikodym derivative is

$$\frac{d\mathcal{P}_{n\varepsilon,m}}{d\mathcal{P}_m}(N) = \exp\{\Sigma_{\omega\in\Omega}[N(D_\omega)\ln\frac{n\mathbb{P}(C_\omega(\varepsilon))}{\bar{c}\,\mathbb{Q}(D_\omega)} - n\mathbb{P}(C_\omega(\varepsilon)) + \bar{c}\,\mathbb{Q}(D_\omega)]\}$$

and let $N(D_\omega)\ln\mathbb{Q}(D_\omega) = 0$ if $\mathbb{Q}(D_\omega) = 0$. From (11) it can be easily deduced that (12) converges to 0: in addition to (11) it is sufficient to notice that

$$|\exp\{\Sigma_{\omega\in\Omega}N(D_\omega)\ln\frac{n\mathbb{P}(C_\omega(\varepsilon))}{\bar{c}\,\mathbb{Q}(D_\omega)}\} - 1| \le \exp\{\Sigma_{\omega\in\Omega}N(D_\omega)|\ln\frac{n\mathbb{P}(C_\omega(\varepsilon))}{\bar{c}\,\mathbb{Q}(D_\omega)}|\} - 1$$

and take the expected value. $\square$

Let $\psi(x), x \in \mathbb{R}^\infty$, be Borel measurable functional in $\mathbb{R}^\infty$ and consider random variables (or *statistics*) $\psi(N_{n\varepsilon})$ and $\psi(N)$, based on countably many $N_n(A_{\gamma_j})$ and $N(B_{\gamma_j})$, respectively. Suppose ψ is such that for any $\delta, \eta > 0$ there exists $m = m(\delta, \eta)$ and functional $\psi_m(x)$ which depends on the first m coordinates of x, such that

$$P\{|\psi(N_{n\varepsilon}) - \psi_m(N_{n\varepsilon})| > \eta\} < \delta \text{ and } P\{|\psi(N) - \psi_m(N)| > \eta\} < \delta\,. \tag{13}$$

Theorem 12 implies that for any such functional we have convergence in distribution

$$\psi(N_{n\varepsilon}) \xrightarrow{d} \psi(N)$$

(see, e.g., [25], Ch.II, §10). The class of functionals satisfying (13) can be relatively wide. It certainly includes statistics of the form

$$\sum_{j=1}^{\infty} c_j g(N_n(A_{\gamma_j}), n\mathbb{P}(A_{\gamma_j})) \text{ with } \sum_{j=1}^{\infty} |c_j| < \infty$$

as soon as

$$E|g(N_n(A_{\gamma_j}), n\mathbb{P}(A_{\gamma_j}))| \to E|g(N(B_{\gamma_j}, \bar{c}\,\mathbb{Q}(B_{\gamma_j})| \le c < \infty,$$

but also may include the so-called Kolmogorov - Smirnov statistic

$$\max_j |N_n(A_{\gamma_j}) - n\mathbb{P}(A_{\gamma_j})| \,,$$

as soon as the class $\{A_{\gamma_j}(\varepsilon), j = 1, 2, \dots\}$ is "appropriately" totally bounded with respect to the semi-metric $\mathbb{P}(\cdot \triangle \cdot)$. Assumption of the total boundedness is commonly used in the theory of weak convergence of empirical processes – see, e.g., [27]. We present it in full detail, for the case of Gaussian convergence of the local process $(N_n, \mathcal{A}(\varepsilon))$, in [11].

References

[1] Z. Artstein. A calculus of set-valued maps and set-valued evolution equations, *Set-Valued Anal.*, **3**(1995), 213-261

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

[3] J.-P. Aubin, H. Frankowska. *Set-valued analysis*, Birkhäuser, 1990

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

[5] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, J. Wellner. *Efficient and Adaptive Estimation in Semiparametric Models*, Johns Hopkins University Press, Baltimore and London *et al.*, 1993.

[6] J.M. Borwein, Q.J. Zhu. A survey of subdifferential calculus with applications. *Nonlinear Analysis*, **38** (1999), 687-773.

[7] E. Carlstein, C. Krishnamoorthy. Boundary estimation *J. Amer. Statist. Assoc.* **87** (1992),430-438.

[8] P. Deheuvels, D.M. Mason. Functional laws of the iterated logarithm for local empirical processes indexed by sets, *Ann. Probab.*, **22** (1994), 1619–1661.

[9] P. Deheuvels, D.M. Mason. Nonstandard local empirical processes indexed by sets, *J. Statist. Plann. Inference*, **45** (1995), 91–112.

[10] J.H.J. Einmahl. Poisson and Gaussian approximation of weighted local empirical processes, *Stochastic Process. Appl.*, **70** (1997), 31–58.

[11] J.H.J. Einmahl, E.Khmaladze. Gaussian limit theorem for local empirical processes in the neighbourhood of a convex body, Working paper, Victoria University of Wellington, 2006.

[12] H. Federer. Curvature measures, *Trans. Amer. Math. Soc.*, **93** (1959), 418–491.

[13] S. Gautier. Affine and eclipsing multifunctions, *Numer. Funct. Anal. Optim.*, **11**(1990), 679-699

[14] D. Hug, G. Last, W. Weil, A local Steiner-type formula for general closed sets and applications, *Math. Z.* **246** (2004), 237–272.

[15] I.A. Ibragimov, R.Z. Has'minskii, *Statistical Estimation. Asymptotic Theory*, Springer, New York, 1981.

[16] E. Khmaladze. Goodness of fit tests for "chimeric" alternatives, *Statist. Neerl.*, **52** (1998), 90-111.

[17] E. Khmaladze. Differentiation of sets in measure, 2005, Report 4-05, School MSCS, Victoria University of Wellington.

[18] E. Khmaladze, and W. Weil. Local empirical processes near boundaries of convex bodies. 2006, accepted to *Ann. Inst. Statist. Mathemat.*

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

[20] C. Lemaréchal and J. Zowe. The eclipsing concept to approximate a multi-valued mapping, *Optimization*, **22** (1991), 3-37

[21] H.-G. Müller and U. Stadtmüller. Discontinuous versus smooth regression, *Ann. Statist.*, **27** (1999), 299-337.

[22] G.Ch. Pflug. *Optimization of Stochastic Models*, Kluwer Acad.Publ., Dordrecht, 1996.

[23] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*, Springer, 1998.

[24] R. Schneider, *Convex Bodies: the Brunn-Minkowski Theory*, Encyclopedia of Mathematics and its Applications, **44**, Cambridge University Press, Cambridge, 1993.

[25] A. Shiryaev, *Probability*, Springer Verlag, 2004

[26] D. Silin. On set-valued differentiation and integration.*Set-Valued Analysis*, **5** (1997),107-146.

[27] A. van der Vaart, J. A. Wellner, *Weak Convergence of Empirical processes*, Springer Verlag, 1996

[28] H. Weisshaupt. A Measure-Valued Approach to Convex Set-Valued Dynamics, *Set-Valued Analysis*, **9**, (2001), 337- 373

Estáte Khmaladze