Petr Lachout

A general version of delta theorem

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 35 (1994), No. 2, 51--57

A General Version of Delta Theorem

P. LACHOUT

Praha*)

Received 15 March 1994

The paper presents a general version of Delta Theorem which helps us to transform probabilistic asymptotic laws. The result is based on famous Topsøe's setup (1972) which is shortly recalled in Introduction. The introduced Delta Theorem for multifunctions improves that of King (1989). In particular, the contingent derivative may be a compact-valued instead of a single-valued multifunction and values are allowed to belong into a Hausdorff linear topological space.

1. Introduction and Topsøe's result

Probability theory has derived a number of weak convergence theorems. From a practical point of view, it is useful to know how these theorems look after a transformation by a given mapping. A well-known result on differentiable function is traditionally called Delta Theorem (or Delta Method).

Delta Theorem. Let ξ_n, ξ be real random variables $\hat{x} \in \mathscr{R}$, $\tau_n > 0$, $t_n \to +\infty$ such that $\tau_n(\xi_n - \hat{x}) \xrightarrow{D} \xi$. Then for any real function f differentiable at the point $\hat{x}$, one can derive $\tau_n(f(\xi_n) - f(\hat{x})) \xrightarrow{D} f'(\hat{x})\xi$.

But the theorem does not cover all requirements. There are problems needing essentially rich spaces. And moreover, the transforming mapping f can be non differentiable and can take several values. For example, this happens when optimization programs are studied.

The crucial question of the setup is how to transform a random variable, generally a probability measure, by a given multifunction at the place of f. An effective construction is due to Topsøe. Let us briefly describe his idea. The following results are stated without any proof, the interested readers can see at Topsøe for them.

A mapping ascribing non-empty subsets of a set Y to points of a set X is called a multifunction. The used notation is $\varphi : X \rightrightarrows Y$. Since a value of φ may contain

*) Institute of Information Theory and Automation, Czech Academy of Sciences, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic

several points, there are two different kinds of inverse images, the strong inverse and the weak inverse. These are given by

$$\varphi^s B = \{x \in X : \varphi(x) \subset B\}, \qquad \varphi^w B = \{x \in X : \varphi(x) \cap B \neq \emptyset\}$$

for any subset B of Y. Always $\varphi^s B \subset \varphi^w B$ and both inverses coincide only if φ is a function, i.e. all values of φ are single-point sets.

In the sequel X and Y are assumed to be Hausdorff spaces because some continuity properties of multifunctions are considered. Following a standard notation, $\mathscr{G}(X)$ (resp. $\mathscr{F}(X)$, $\mathscr{K}(X)$, $\mathscr{B}(X)$) denotes the collection of all open (resp. closed, compact, Borel) sets in X.

Definition 1. A multifunction $\varphi : X \rightrightarrows Y$ is said to be

(i) upper semicontinous (u.s.c.) if the strong inverse of any open subset of Y is open in X;

(ii) lower semicontinuous (l.s.c.) if the weak inverse of any open subset of Y is open in X;

(iii) continuous if it is both u.s.c. and l.s.c.;

(iv) closed-valued if its values are closed in Y;

(v) compact-valued if its values are compact in Y.

There are some equivalent definitions employing general nets (see Topsøe for the proof).

Lemma 1. *Let X, Y be Hausdorff spaces and $\varphi : X \rightrightarrows Y$ be a multifunction. Then*

(i) *φ is l.s.c. $\Leftrightarrow$ if $x_\alpha \in X$, $x_\alpha \to x \in X$ and $y \in \varphi(x)$ then there exist a subnet x_β and a choice $y_\beta \in \varphi(x_\beta)$ such that $y_\beta \to y$;*

(ii) *φ is compact-valued u.s.c. $\Leftrightarrow$ if a net $x_\alpha \in X$, $x_\alpha \to x \in X$, $y_\alpha \in \varphi(x_\alpha)$ then there exist a subnet such that $y_\beta \to y \in \varphi(x)$.*

In this case we say that φ preserves compact nets.

The action of a multifunction can be extended to probability measures. Topsøe considered a general case, but for our purpose Radon probability measures are sufficient. The set of all Radon probability measures on a Hausdorff space X will be denoted by $\mathscr{P}(X)$; i.e. $\mu \in \mathscr{P}(X)$ provided that $\mu(X) = 1$ and

$$\mu(B) = \sup\{\mu(K) : B \supset K \in \mathscr{K}(X)\} \quad \text{for each} \quad B \in \mathscr{B}(X).$$

Definition 2. Let X, Y be Hausdorff spaces and $\varphi : X \rightrightarrows Y$ be a multifunction. We define the mapping $\tilde{\varphi} : \mathscr{P}(X) \to \exp \mathscr{P}(Y)$ by $\nu \in \tilde{\varphi}(\mu) \Leftrightarrow \nu^*(B) \geqq \mu^*(\varphi^s B)$ for every $B \subset Y$.

Recall that the upper star indicates the outer measure and the lower star indicates the inner measure. The extension can be equivalently expressed in several ways (see Topsøe for the proof).

Lemma 2. *Let X, Y be Hausdorff spaces and $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$. Then the following conditions are equivalent.*

(i) $\nu \in \tilde{\varphi}(\mu)$;
(ii) $\mu^*(\varphi^s B) \leqq \nu^*(B)$ *for every* $B \subset Y$;
(iii) $\mu_*(\varphi^w B) \geqq \nu_*(B)$ *for every* $B \subset Y$;
(iv) $\mu^*(A) \leqq \nu^*(\varphi A)$ *for every* $A \subset X$;
(v) $\mu^*(\varphi^s G) \leqq \nu(G)$ *for every* $G \in \mathscr{G}(Y)$;
(vi) $\mu(K) \leqq \nu^*(\varphi K)$ *for every* $K \in \mathscr{K}(X)$.

The introduced mapping is not necessarily a multifunction because $\tilde{\varphi}(\mu)$ may be empty for some $\mu \in \mathscr{P}(X)$. Nevertheless, a few nice properties are held (see Topsøe for the proofs).

Corollary 1. *For each* $x \in X$, *one has*

$$\nu \in \tilde{\varphi}(\delta_x) \Leftrightarrow \nu^*(\varphi(x)) = 1\,,$$

where δ_x *denotes the probability measure concentrated at the point x.*

Especially, $\tilde{\varphi}(\delta_x) \neq \emptyset$ since $\delta_y \in \tilde{\varphi}(\delta_x)$ for every $y \in \varphi(x)$ and $\tilde{\varphi}(\mu) \neq \emptyset$ if the support of μ consists of a finite number of points.

Corollary 2. *Let* $\mu \in \mathscr{P}(X)$ *be such that the multifunction* φ *is a* μ*-a. s. function, i.e.* $\mu^*(x \in X : \#\varphi(x) \geqq 2) = 0$. *Then* $\tilde{\varphi}(\mu)$ *contains at most one point. (*$\#A$ *denotes the cardinality of the set A)*

Corollary 3. *Let X, Y be Polish spaces and* ξ *be a random variable with the probability distribution* μ. *If* $\zeta \in \varphi(\xi)$ *is a measurable selection. i.e.* ζ *is a random variable and* $\zeta(\omega) \in \varphi(\xi(\omega))$, *with the probability distribution* ν *then* $\nu \in \tilde{\varphi}(\mu)$ *(in this case, every probability measure is necessarily Radon).*

The reverse question if every $\nu \in \tilde{\varphi}(\mu)$ can be represented by a measurable selection $\zeta \in \varphi(\xi)$ is difficult. The answer depends on the structure of the probability space in question as well as of the Polish spaces X, Y.

The constructed mapping saves the upper semicontinuity, as Topsøe has shown.

Proposition (Topsøe). *Let X, Y be Hausdorff spaces and* $\varphi : X \rightrightarrows Y$ *be a compact-valued u.s.c. multifunction. Then* $\tilde{\varphi} : \mathscr{P}(X) \rightrightarrows \mathscr{P}(Y)$ *is a compact-valued u.s.c. multifunction as well.*

For the proof see the theorem 3.13 of Topsøe.

2. Delta Theorem

The presented setup gives an effective tool of investigation of a general Delta Theorem employing multifunctions. For that purpose, a notion of multifunction differentiability needs an explanation. But this requires a special topological structure. Therefore in this chapter the spaces X, Y are assumed to be Hausdorff linear topological spaces.

There are several definitions of a derivative of a multifunction. All of them are constructed by means of special tangent cones of the multifunction graph, see Aubin & Frankowska. Our result needs the contingent derivative, cf. the definition 5.1.1 of Aubin & Frankowska. Moreover, the considered multifunction has to fulfil a uniformity condition.

Definition 3. Let X, Y be Hausdorff linear topological spaces. A multifunction $\varphi : X \rightrightarrows Y$ is said to fulfil the property (L) at the point $\hat{x} \in X$ if

(L) there is $T : X \rightrightarrows Y$ a compact-valued u.s.c. cone multifunction, i.e. $T(\alpha y) = \alpha T(y)$ for every $\alpha \geqq 0$, $y \in X$, such that $\varphi(x + \hat{x}) \subset \varphi(\hat{x}) + T(x)$ on a neighbourhood of the point zero.

If Y is a finite-dimensional Euclidean space then the property (L) is equivalent to the local upper Lipschitz property, for definition see Aubin & Frankowska or King. The property (L) allows an equivalent definition of the contingent derivative.

Lemma 3. *Let X, Y be Hausdorff linear topological spaces, $\hat{x} \in X$, $\varphi : X \rightrightarrows Y$ be a multifunction fulfilling the property (L) at the point $\hat{x}$ and $\varphi(\hat{x}) = \{\hat{y}\}$. Then there exists the multifunction $D : X \rightrightarrows Y$ uniquely determined by the following two conditions*

(i) *If $t_\alpha > 0$, $x_\alpha \in X$, $y_\alpha \in \varphi(t_\alpha x_\alpha + \hat{x})$, $t_\alpha \to 0$, $x_\alpha \to x \in X$ then there exists a subnet such that $\frac{1}{t_\beta}(y_\beta - \hat{y}) \to y \in D(x)$.*

(ii) *If $x \in X$, $y \in D(x)$ then there exists a net $t_\alpha > 0$, $x_\alpha \in X$, $y_\alpha \in \varphi(t_\alpha x_\alpha + \hat{x})$ such that $t_\alpha \to 0$, $x_\alpha \to x$ and $\frac{1}{t_\alpha}(y_\alpha - \hat{y}) \to y$.*

The multifunction D is called the contingent derivative of φ at the point $(\hat{x}, \hat{y})$ and is a compact-valued u.s.c. cone multifunction, i.e. $D(\alpha x) = \alpha D(x)$ for every $x \in X$, $\alpha \geqq 0$.

Proof. Evidently, there is at most one multifunction fulfilling (i) and (ii).

Define a mapping $D : X \to \exp Y$ by

$$y \in D(x) \Leftrightarrow \text{there exist a net } t_\alpha > 0,\ x_\alpha \in X,\ y_\alpha \in \varphi(t_\alpha x_\alpha + \hat{x}) \text{ such that } t_\alpha \to 0,\ x_\alpha \to x,\ \frac{1}{t_\alpha}(y_\alpha - \hat{y}) \to y\,.$$

a) We have to show that D is a multifunction, i.e. $D(x)$ should be always a non-empty set.

Let $x \in X$. Take the sequence $t_n = \frac{1}{n}$, $x_n = x$, $y_n \in \varphi(t_n x + \hat{x})$. Hence, $y_n \in \hat{y} + T(t_n x)$ for n large enough, because φ fulfils the property (L) at the point $\hat{x}$. Since T is a compact-valued u.s.c. cone multifunction and

$\frac{1}{t_n}(y_n - \hat{y}) \in T(x)$, there exists a subnet $\frac{1}{t_\beta}(y_\beta - \hat{y}) \to T(x)$. Consequently, $y \in D(x)$ and thus D is a multifunction.

b) We have to verify the property (i) since (ii) holds immediately.
Let $t_\alpha > 0, x_\alpha \in X, y_\alpha \in \varphi(t_\alpha x_\alpha + \hat{x}), t_\alpha \to 0, x_\alpha \to x \in X$. Hence, $y_\alpha \in \hat{y} + T(t_\alpha x_\alpha)$ for α large enough, because φ fulfils the property (L) at $\hat{x}$. Since T is a compact-valued u.s.c. cone multifunction $\frac{1}{t_\alpha}(y_\alpha - \hat{y}) \in T(x_\alpha)$ and there exists a subnet $\frac{1}{t_\beta}(y_\beta - \hat{y}) \to y \in T(x)$. Consequently, $y \in D(x)$.

c) We have to prove that D is a compact-valued u.s.c. cone multifunction.
Evidently, D is a cone multifunction and $D(x) \subset T(x)$ for every $x \in X$.
Let $x_\alpha \in X$, $x_\alpha \to x \in X$, $y_\alpha \in D(x_\alpha)$. There exists a convergent subnet $y_\beta \to y \in T(x)$ since $y_\alpha \in D(x_\alpha) \subset T(x_\alpha)$ and T is compact-valued u.s.c. The spaces X, Y are Hausdorff linear topological spaces, therefore one can find another net $t_\gamma > 0$, $x'_\gamma \in X$, $y'_\gamma \in \varphi(t_\gamma x'_\gamma + \hat{x})$ such that $t_\gamma \to 0$, $x'_\gamma \to x$, $\frac{1}{t_\gamma}(y'_\gamma - \hat{y}) \to y$. Consequently, $y \in D(x)$ and D is compact-valued u.s.c.

Q.E.D.

Theorem. *Let X, Y be Hausdorff linear topological spaces, $\hat{x} \in X$, $\varphi : X \rightrightarrows Y$ be a compact-valued u.s.c. multifunction fulfilling the property (L) at the point $\hat{x}$ and $\varphi(\hat{x}) = \{\hat{y}\}$. Let ξ_α be a net of random variables with values in X and $\tau_\alpha(\xi_\alpha - \hat{x}) \xrightarrow{\mathscr{D}} \xi$ for some standardization $\tau_\alpha > 0$, $\tau_\alpha \to +\infty$ and $\zeta_\alpha \in \varphi(\xi_\alpha)$ be measurable selections. Assume that the probability measures induced by ξ_α, ξ and ζ_α are Radon probability measures. Then there exists a subnet such that $\tau_\beta(\zeta_\beta - \hat{y}) \xrightarrow{\mathscr{D}} \zeta$ and the probability measures μ of ξ and ν of ζ satisfy $\nu \in \tilde{D}_\varphi(\mu)$, where D_φ is the contingent derivative of φ at the point $(\hat{x}, \hat{y})$.*

Unfortunately, it is not sure that ζ can be found as a measurable selection of $D_\varphi(\xi)$.

Proof. Consider the multifunction $\psi : R_+ \times X \rightrightarrows R_+ \times Y$ given by

$$\psi(t, x) = \left(t, \frac{1}{t}(\varphi(tx + \hat{x}) - \hat{y})\right) \quad \text{if} \quad t > 0, x \in X,$$

$$\psi(0, x) = (0, D_\varphi(x)) \quad \text{if} \quad x \in X.$$

Our task is to show that ψ is a compact-valued u.s.c. multifunction. The assertion of the theorem immediately follows this fact.

It is sufficient to verify that ψ preserves compact nets. Let $t_\alpha \geqq 0$, $x_\alpha \in X$, $t_\alpha \to t \geqq 0$, $x_\alpha \to x \in X$ and $y_\alpha \in Y$ such that $(t_\alpha, y_\alpha) \in \psi(t_\alpha, x_\alpha)$. We are looking for a convergent subnet $y_\beta \to y \in Y$ and $(t, y) \in \psi(t, x)$. There are three different cases.

a) Suppose $t > 0$.

Hence, $y_\alpha \in \frac{1}{t_\alpha}\left(\varphi(t_\alpha x_\alpha + \hat{x}) - \hat{y}\right)$ and therefore $t_\alpha y_\alpha + \hat{y} \in \varphi(t_\alpha x_\alpha + \hat{x})$. Since $t_\alpha x_\alpha \to tx$ and φ is compact-valued u.s.c., there exists a subnet such that $t_\beta y_\beta + \hat{y} \to ty + \hat{y} \in \varphi(tx + \hat{x})$. Consequently, $y_\beta \to y$ and $(t, y) \in \psi(t, x)$.

b) Suppose $t_\gamma = 0$ for a subnet.
Hence, $y_\gamma \in D_\varphi(x_\gamma)$ and $x_\gamma \to x$. D_γ is a closed-valued u.s.c. multifunction according to Lemma 3. Therefore, one can select a convergent subnet $y_\beta \to y \in D_\varphi(x)$ and $(0, y) \in \psi(0, x)$.

c) Suppose $t = 0$ and $t_\alpha > 0$ for all α large enough.
Hence, $y_\alpha \in \frac{1}{t_\alpha}\left(\varphi(t_\alpha x_\alpha + \hat{x}) - \hat{y}\right)$ if α is large enough and $t_\alpha \to 0$, $x_\alpha \to x$. Since D_φ is the contingent derivative of φ, there exists a convergent subnet $y_\beta \to y \in D_\varphi(x)$ and $(0, y) \in \psi(0, x)$.

We have verified that ψ is a compact-valued u.s.c. multifunction. This concludes the proof as follows.

We have

$$\left(\frac{1}{\tau_\alpha}, \tau_\alpha(\zeta_\alpha - \hat{y})\right) \in \psi\left(\frac{1}{\tau_\alpha}, \tau_\alpha(\xi_\alpha - \hat{x})\right).$$

Denote by μ'_α (resp. ν'_α) the probability distribution of $\tau_\alpha(\xi_\alpha - \hat{x})$ (resp. $\tau_\alpha(\zeta_\alpha - \hat{y})$). These measures are evidently Radon measures because ξ_α, ζ_α are assumed to induce Radon measures. Hence, we derive $\delta_{\frac{1}{\tau_\alpha}} \otimes \nu'_\alpha \in \tilde{\psi}\,(\delta_{\frac{1}{\tau_\alpha}} \otimes \mu'_\alpha)$. Since $\tilde{\psi}$ is compact-valued u.s.c., there exist a subnet

$$\delta_{\frac{1}{\tau_\beta}} \otimes \nu'_\beta \xrightarrow{\omega} \varrho \in \tilde{\psi}(\delta_0 \otimes \mu),$$

where μ denotes the probability distribution of ξ (which is Radon, too). From the definition of $\tilde{\psi}$ we obtain

$$\varrho(\{0\} \times Y) \geqq (\delta_0 \otimes \mu)^*(\psi^s(\{0\} \times Y)) = \delta_0 \otimes \mu(\{0\} \times X) = 1.$$

Hence, $\varrho = \delta_0 \otimes \nu$ for some probability measure $\nu \in \mathscr{P}(Y)$. Moreover, $\nu(G) = \varrho(\{0\} \times G) \geqq (\delta_0 \otimes \mu)^*(\psi^s(\{0\} \times G)) = \delta_0 \otimes \mu(\{0\} \times D^s_\varphi G) = \mu(D^s_\varphi G)$ for every $G \in \mathscr{G}(Y)$. This gives $\nu \in \tilde{D}_\varphi(\mu)$ according to Lemma 2. If $G \in \mathscr{G}(Y)$, then $R_+ \times G$ is open in $R_+ \times Y$ and thus the weak convergence yields

$$\liminf_\beta \nu'_\beta(G) = \liminf_\beta \delta_{\frac{1}{\tau_\beta}} \otimes \nu'_\beta(R_+ \times G) \geqq \varrho(R_+ \times G) = \nu(G),$$

which is $\nu'_\beta \xrightarrow{\omega} \nu \in \tilde{D}_\varphi(\mu)$. Consequently, $\frac{1}{\tau_\beta}(\zeta_\beta - \hat{y}) \xrightarrow{\mathscr{D}} \zeta$ where the probability distribution of ζ is the measure ν, note that such random variable always exists.

The above theorem improves the theorem 4.3 of King. In particular, the target space Y is allowed to be a Hausdorff linear topological space instead of a finite-dimensional Euclidean space and the assumption that the contingent derivative is a.s. single-valued is removed, too. But one cannot expect any better answer than a convergent subnet, in such a general case.

References

[1] AUBIN, J.-P. AND FRANKOWKA, H., Set-Valued Analysis (Birkhäuser, Boston 1990).

[2] KING, A J., Generalized Delta Theorem for multivated mappings and measurable selections, Math. Oper. Res. 14, 4 (1989) 720–736.

[3] LACHOUT, P., On multifunction transforms of probability measures (submitted to the Transactions of the Sixth International Conference on Stochastic Programming, 1992)

[4] TOPSØE, F., Measurable spaces connected by correspondence, Math. Scand. 30 (1972) 5–45.

[5] KELLEY, J. L., General Topology, (D. van Nostrand Comp., New York 1955).