# MascotNum2014 conference - Multivariate Quantile Surfaces and Application to an Aircraft Problem

Adil Ahidar
*Université Toulouse III - Paul Sabatier*

**Affiliation**: - *Institut de Mathématiques de Toulouse (IMT)*, Université Paul Sabatier 118, route de Narbonne 31062 Toulouse, France
- *Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS)*, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France

**Email**: adil.ahidar@math.univ-toulouse.fr – **URL**: http://www.math.univ-toulouse.fr

**Master:** Université Toulouse III - Paul Sabatier

**Ph.D.** (2011-2014): Université Toulouse III - Paul Sabatier

**Supervisor(s):** Prof. Philippe Berthet (Université Toulouse III) and Dr. Thierry Klein (Université Toulouse III) and Prof. Serge Gratton (CERFACS) and Dr. Jayant Sengupta (EADS IW)

**Abstract:**

**Introduction** During an aircraft design process, an engineer would be interested to distribute margins on different suppliers in an optimal way, i.e. giving more margins to those who do not control their own design process (because of a new technology for instance) and less margins to those who can easily or cheaply commit on a good confidence. The design, influenced by the uncertainty of the suppliers margins, has to comply to some performance constraints taking into account a risk measure. In the case of multiple criteria, this leads to the definition of a multivariate quantile. The main contribution of this PhD thesis is on a new definition and on the associated convergence theroem of a multivariate quantile.

**Multivariate Quantile** Let $X_1, ..., X_n$ be i.i.d. $\mathbb{R}^d$ valued random variables with common law $P = \mathbb{P}^X$. Many approaches in multivariate data analysis have been proposed to describe the structure of such a data cloud. In connection with our work it is worth mentioning spatial quantiles, data depth, level sets, mode localization, shorth sets, classification, $k$-means, trimming, among others.

In this presentation we introduce an original notion of multivariate quantiles that is not based on a global $M$-estimation but rather on a directional $M$-estimation which is also a natural generalization of the univariate quantiles.

For $d = 1$, let's first recall the definition of the univariate quantile $Q(\alpha) = \inf\{y \in \mathbb{R} : \mathbb{P}(X \leq y) \geq \alpha\}$

Facing the usual fact that $\mathbb{R}^d$ is not ordered our idea is simply to admit subjectivity and thus to define a local viewpoint rather than a global one, anchored at some point of reference $O$ and arbitrary shape $\varphi$ with the motivation of crossing information gathered by changing viewpoint $O$, shape $\varphi$ and $\alpha$-th order of quantile. Since these viewpoints are correlated, the next key steps will be to properly compare them to automatically learn about $P$, but this is beyond our current scope. In this presentation we mainly focus on the special case of half-spaces instead of the general shape $\varphi$.

For $d \geq 1$, we define the half-space standing at distance $y \in \mathbb{R}$ from $O \in \mathbb{R}^d$ in the direction $u \in \mathbb{S}_{d-1}$

$$H(O,\, u,\, y) = \left\{x \in \mathbb{R}^d : \langle x - O,\, u\rangle \leq y\right\}$$

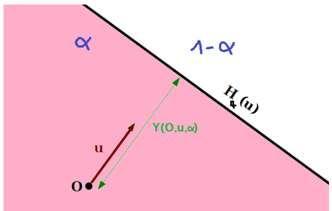
now, given $\alpha \in (1/2, 1]$, we define the $\alpha$-th quantile range from $O$ in the direction $u$

$$Y(O, u, \alpha) = \inf\{y \in \mathbb{R} : \mathbb{P}(H(O,\, u,\, y)) \geq \alpha\}$$

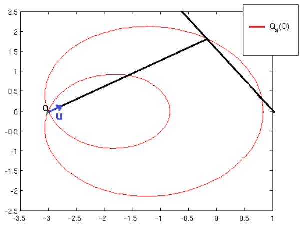
and the $\alpha$-th quantile point seen from $O$ in the direction $u$

$$Q(O, u, \alpha) = O + Y(O, u, \alpha) \cdot u$$

Imagine an observer located in $O \in \mathbb{R}^d$ looking at the sample in all directions $u \in \mathbb{S}_{d-1}$ where $\mathbb{S}_{d-1}$ is the unit sphere of $\mathbb{R}^d$. How can he picture out the empirical mass localization in $\mathbb{R}^d$ ? We propose to




(a) in pink: space region of mass $\alpha$
in black: half-space H(O,y,$\alpha$)



(b) in red: The quantile surface
$Q(O, \alpha) = \{Q(O, u, \alpha) : \mathbb{S}_{d-1}\}$

draw a collection of $u$-directional $\alpha$-th quantile points that we call a subjective $\alpha$-th quantile surface, for a fixed $\alpha \in (1/2, 1)$.

Each point of this surface (b) is the univariate $\alpha$-th quantile of the projected distribution on the line $(O, u)$. How can the observer catch more information on the law $P$ Keeping $O$ fixed and letting $\alpha$ vary determines $P$. Thus analyzing a collection of such surfaces can be viewed as some purely non-parametric and non-linear data analysis approach in which we keep track of more than a few orthogonal directions. A second motivation for these $\alpha$-th quantile surfaces is that the algorithms computing them at a reasonable precision are simple and fast, even for large samples, faster than level sets estimators or plug-in estimators based on density estimators – however density may even not exist.

**Convergence Results** The graphical representations of these spatial quantiles are random closed surfaces generated by the sample for which we establish the following theorems:

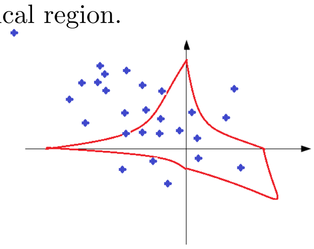
1-Almost sure consistency // 2-Uniform central limit theorem with some rate // 3-Uniform Strong approximation with rate // 4-Bahadur-Kiefer type representation

**Back to the industrial problem** The industrial problem can be reformulated (or generalized) in the folowing way:

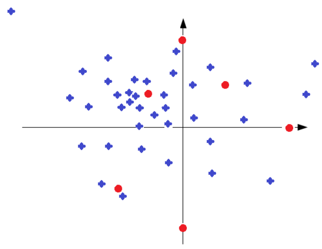
$$\begin{cases} \max_{\sigma \in \mathbb{R}^n} C(\sigma) \\ s.t.\, Z_\sigma^\alpha \subset \mathcal{M} \end{cases}$$

$C : \sigma \mapsto C(\sigma)$ is a controlability function and $Z_\sigma^\alpha$ some transformation of our quantile of level $\alpha$ and the set $\mathcal{M}$ can be seen as a confidence region

In other words, on the space of the outputs some regions are critical (in a weaker version we only consider some critical points ) and one wants to calibrate $\sigma$ such that with high probability the outputs are far from the critical region.



(a) in red: boundary of critical region
in blue: Sample points



(b) in red: critical points
in blue: Sample points

We show that the quantiles introduced in the previous part are well-suited to handle with this type of applied problems.

**Bibliography**:

[1] D. Nolan, Asymptotics for multivariate trimming, Stochastic Processes and their Applications, Volume 42, Issue 1, August 1992, Pages 157-169

[2] J. Tukey, Mathematics and the picturing of data, In Proc. 1975 Inter. Cong. Math., Vancouver 523531. Montreal: Canad. Math. Congress.

**Short biography** – The PhD is a follow-up to an internship completed on a landing gear problem in EADS IW in 2010/11. The PhD is founded at "CERFACS" by "EADS Innovation Works".