

# Application de la classification à vecteurs de support pour l'estimation de quantiles multidimensionnels extrêmes

Miguel Piera Martínez<sup>1</sup> \*, Emmanuel Vazquez<sup>1</sup>,  
Gilles Fleury<sup>1</sup>, Eric Walter<sup>2</sup>

1. Département Signaux et Systèmes Électroniques, Supélec, Gif-sur-Yvette, France

2. L2S, CNRS - Supélec - Univ. Paris-Sud, Gif-sur-Yvette, France

## Résumé

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$  et de loi de probabilité  $P$ . L'objectif de cet article est de déterminer empiriquement à partir d'un  $n$ -échantillon de  $\mathbf{X}$  un ensemble  $Q_{1-\alpha} \subset \mathbb{R}^d$ , appelé quantile multidimensionnel, de  $P$ -probabilité  $1 - \alpha$  avec  $\alpha$  petit, notamment  $\alpha < 1/n$ . La méthode utilisée est fondée sur une méthode de classification à vecteurs de support [1] que nous utilisons conjointement avec la théorie des valeurs extrêmes afin d'extrapoler le comportement des données observées, c'est-à-dire lorsque l'on cherche  $Q_{1-\alpha}$  avec  $\alpha < 1/n$ .

*Mots clés : analyse des données, théorie des valeurs extrêmes, méthodes à vecteurs de support, méthodes à noyaux, quantiles multidimensionnels de volume minimal.*

Let  $\mathbf{X}$  be a  $d$ -dimensional random vector with probability law  $P$ . This paper deals with the estimation from an  $n$ -sample of a multivariate quantile set  $Q_{1-\alpha} \subset \mathbb{R}^d$  of  $P$ -probability  $1 - \alpha$ , with  $\alpha$  small — in particular when  $\alpha < 1/n$ . The method is based on support vector classification [1] together with extreme value theory in order to extrapolate the behaviour of the data, i.e, we search for  $Q_{1-\alpha}$  with  $\alpha < 1/n$ .

## 1 Introduction

La complexité d'un système et les approximations de modélisation qui en résultent, le caractère aléatoire des perturbations externes ainsi que la dispersion des paramètres de conception autour de leur valeur nominale sont autant de raisons qui amènent à remettre en cause les approches déterministes supposant une connaissance parfaite du système et de son environnement. La nécessité de concevoir des systèmes robustes nous conduit à élaborer des modèles statistiques qui permettent de quantifier l'effet des incertitudes, et en particulier l'apparition de valeurs extrêmes à la sortie des systèmes. En général, les valeurs extrêmes apparaissent avec une faible probabilité, ce qui rend leur évaluation particulièrement difficile.

Dans cet article, nous proposons une approche permettant de déterminer des ensembles de faible probabilité lorsque les observations du système ou du phénomène physique étudié sont

---

\*soutenu par la Fondation d'entreprise EADS ([www.fondation.eads.net](http://www.fondation.eads.net))

multivariées. Cette approche proposée repose sur une classification de type *SVM à une classe*, expliquée brièvement dans la section 3, et l'utilisation de la *théorie des valeurs extrêmes*, dont nous rappelons un théorème fondamental dans la section 4. Nous illustrons la méthode dans la section 5.

## 2 Quantiles multidimensionnels

Soient  $(\mathbb{X}, \mathcal{A}, \mathbb{P})$  un espace de probabilités,  $\mathcal{Q}$  une classe de sous-ensembles  $\mathbb{P}$ -mesurables et  $\lambda$  une fonction à valeurs réelles définie sur  $\mathcal{Q}$ . Un *quantile multidimensionnel* par rapport à  $(\mathbb{X}, \mathcal{A}, \mathbb{P}, \mathcal{Q}, \lambda)$  est un ensemble  $Q_{1-\alpha}$ , non nécessairement unique, qui atteint l'infimum

$$u(\alpha) = \inf\{\lambda(Q) : Q \in \mathcal{Q}, \mathbb{P}(Q) \geq 1 - \alpha\}, \quad 0 < \alpha \leq 1.$$

La fonction  $u(\alpha)$  est appelée *fonction quantile* [2]. Si  $\mathcal{Q}$  est une  $\sigma$ -algèbre et  $\lambda$  la mesure de Lebesgue, alors  $Q_{1-\alpha}$  est l'ensemble de volume minimal contenant au moins une fraction  $1 - \alpha$  de la masse de probabilité. Lorsque  $\mathbb{P}$  correspond à une distribution empirique obtenue à partir d'un  $n$ -échantillon d'un vecteur aléatoire  $\mathbf{X} \in \mathbb{X}$ , on obtient un quantile empirique. Observons également que si  $\alpha$  tend vers zéro, l'estimation du quantile devient un problème d'estimation du support d'une densité de probabilité multidimensionnelle.

De nombreux travaux ont été effectués dans le cadre de l'estimation d'ensembles  $Q_{1-\alpha}$  de volume minimal (voir notamment [3, 4, 5, 6]). Si  $\hat{f}_n(x)$  est un estimateur de la densité de probabilité de  $\mathbf{X}$  calculée à partir d'un  $n$ -échantillon, alors  $\hat{Q}_{1-\alpha} = \{x \in \mathbb{X} : \hat{f}_n(x) \leq \alpha\}$  est un quantile empirique de volume minimal dit de type *plug-in* [7].

Dans cet article, nous considérons un  $n$ -échantillon d'une variable aléatoire  $\mathbf{X} \in \mathbb{X}$  de loi  $\mathbb{P}$ , et l'objectif est d'estimer un quantile multidimensionnel dans le cas  $0 \leq \alpha \leq 1/n$  (qui correspond à une extrapolation du comportement des données observées). Par la suite,  $\mathcal{Q} = \{B_\rho, \rho \in \mathbb{R}^+\}$  est une classe de sous-ensembles paramétrés telle que

$$\rho_1 < \rho_2 \Rightarrow B_{\rho_1} \supset B_{\rho_2}, \quad (1)$$

et  $\lambda$  est défini par  $\lambda(B_\rho) \triangleq \rho$ , pour tout  $B_\rho \in \mathcal{Q}$ . Dans la section suivante, nous définissons une telle classe  $\mathcal{Q}$  à l'aide d'une méthode de classification à vecteurs de support, dite à une seule classe [1, 8].

## 3 Classification à vecteurs de support

Un algorithme de classification à *une seule classe* a pour objectif de déterminer un ensemble régulier en un certain sens contenant une fraction fixée des données observées. De nombreux algorithmes sont envisageables. L'un des plus classiques en statistique repose sur l'estimateur de Parzen d'une densité de probabilité. Le seuillage de la densité estimée fournit des ensembles de niveaux vérifiant la propriété (1). Dans cette étude, nous retenons un algorithme de type vecteurs de support (SVM) décrit dans [1]. Cet algorithme fournit (sous certaines conditions) un estimateur consistant d'une densité de probabilité et par conséquent des estimées des quantiles  $Q_{1-\alpha}$  de volume asymptotiquement minimal [8]. Les avantages de cet algorithme sont de

deux types. D'une part, la classe  $\mathcal{Q}$  est décrite à partir d'un petit nombre de données (les vecteurs de support). D'autre part, il s'agit d'un algorithme régularisé (cette propriété est souhaitable dans le cadre de l'estimation de quantiles multidimensionnels).

Soit  $\mathcal{F}$  un espace hilbertien de fonctions réelles définies sur  $\mathbb{X}$ , muni d'un produit scalaire noté  $(\cdot, \cdot)_{\mathcal{F}}$ . Supposons de plus qu'il existe une fonction  $k : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ , appelée *noyau reproduisant*, telle que

$$\forall f \in \mathcal{F}, \quad f(\mathbf{x}) = (f, k(\mathbf{x}, \cdot))_{\mathcal{F}}, \quad \forall \mathbf{x} \in \mathbb{X}.$$

$\mathcal{F}$  est alors un espace à noyau reproduisant [9].

Soient  $\mathbf{x}_i \in \mathbb{R}^d$ ,  $i = 1, \dots, n$ , des données modélisées comme des réalisations d'un vecteur aléatoire  $\mathbf{X}$ . Une méthode de classification à vecteurs de support [1] construit une fonction  $\hat{f}_{\lambda, n} \in \mathcal{F}$  minimisant le critère régularisé

$$\frac{1}{n} \sum_{i=1}^n \max(0, 1 - f(\mathbf{x}_i)) + \lambda \|f\|_{\mathcal{F}}^2, \quad \lambda > 0. \quad (2)$$

Le théorème du représentant [10] montre que  $\hat{f}_{\lambda, n}$  peut se mettre sous la forme

$$\hat{f}_{\lambda, n} : \mathbf{x} \mapsto \sum_{i=1}^n \hat{a}_{\lambda, i} k(\mathbf{x}_i, \mathbf{x}). \quad (3)$$

De plus, il est possible de montrer que  $\hat{f}_{\lambda, n}$  est l'unique solution d'un programme quadratique [1].

Notons  $B_{\rho}^{\lambda, n} = \{\mathbf{x} \in \mathbb{X} : \hat{f}_{\lambda, n}(\mathbf{x}) > \rho\}$  et  $\mathcal{Q}_n$  la famille  $\{B_{\rho}^{\lambda, n}\}_{\rho}$ . [1] montre que  $\mathbb{P}(B_{\rho^*}^{\lambda^*, n})$  tend vers  $1 - \alpha$  lorsque  $n$  tend vers l'infini pour des valeurs  $\rho^*$  et  $\lambda^*$  convenablement choisies. De plus,  $\hat{f}_{\lambda^*, n}(\mathbf{x})$  est un estimateur de la densité de probabilité de  $\mathbf{X}$  tronquée au seuil  $2\lambda^*$  [8].  $B_{\rho^*}^{\lambda^*, n}$  tend alors vers un quantile de volume minimal si  $\rho \leq 2\lambda^*$ . La figure 1 représente des quantiles au niveau  $1 - \alpha = 0.5$  estimés par cette méthode avec  $\lambda^* = 1/n\alpha$  et  $\rho^*$  choisi par la méthode décrite dans [1].

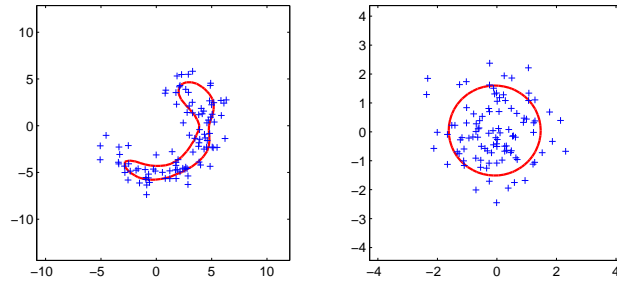


FIG. 1 – Quantile empirique multidimensionnel au niveau  $1 - \alpha = 0.5$  (représenté en trait plein) estimé par classification à vecteurs de support des données (représentées par des croix) pour deux distributions. Le noyau est choisi gaussien avec  $\sigma = 3$ , le choix de  $\sigma$  est également décrit dans [8].

Si l'on cherche un  $\hat{Q}_{1-\alpha} \in \mathcal{Q}_n$  tel que  $\mathbb{P}(\hat{Q}_{1-\alpha}) \approx 1 - \alpha$  avec  $\alpha < 1/n$ , il est possible d'améliorer l'estimation en caractérisant la queue de la loi de probabilité de  $\hat{f}_{\lambda, n}(\mathbf{X})$  à l'aide

de la *théorie des valeurs extrêmes*. L'idée est de choisir dans (2) un  $\lambda$  qui permette d'avoir un nombre convenable de  $\mathbf{x}_i$  pour lesquels  $\widehat{f}_{\lambda,n}(\mathbf{x}_i) < 2\lambda$  et d'estimer  $\rho$  tel que  $\mathbb{P}\{\widehat{f}_{\lambda,n}(\mathbf{X}) < \rho\} = \alpha$  en utilisant le résultat rappelé dans la section suivante.

## 4 Théorie des valeurs extrêmes

La *théorie des valeurs extrêmes* (TVE) est une branche des statistiques développée dans les années 1950 qui fournit des résultats asymptotiques sur le comportement des queues de probabilité [11]. Dans le cas unidimensionnel, un quantile au niveau  $1 - \alpha > 1 - 1/n$ , avec  $n$  le nombre d'observations, est appelé *quantile extrême*. L'estimation d'un quantile extrême est un problème important dans de nombreuses applications (en finance, en ingénierie, en climatologie...). Comment estimer un quantile à 99% si l'on ne dispose que de 50 observations? L'utilisation d'un modèle paramétrique est nécessaire lorsqu'on veut extrapoler le comportement des données observées.

La TVE est fondée sur deux théorèmes fondamentaux asymptotiques : celui de Fisher - Tippett et celui de Pickands - de Haan. Bornons-nous à rappeler le théorème de Pickands - de Haan, qui établit le comportement limite des queues de loi de probabilité de variables aléatoires.

**Théorème** (Pickands - de Haan). *Soient  $X_1, X_2, \dots, X_n$   $n$  variables aléatoires i.i.d. de même fonction de répartition  $F$ , alors*

$$F_u(x) \triangleq \mathbb{P}\{X - u > x | X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad 0 \leq x < x_0 - u, \quad (4)$$

avec

$$x_0 = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty.$$

Si il existe deux suites  $(a_n) \in \mathbb{R}^{+*}$  et  $(b_n) \in \mathbb{R}$  telles que

$$\mathbb{P}\left\{\frac{1}{a_n}(\max(X_1, X_2, \dots, X_n) - b_n) \leq x\right\} = F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x),$$

où  $G(x)$  est une fonction de répartition non dégénérée, alors

$$\lim_{u \rightarrow x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - P_{\xi, \beta(u)}(x)| = 0, \quad (5)$$

où  $P_{\xi, \beta(u)}(x)$  est la distribution de Pareto généralisée (DPG), qui s'exprime sous la forme

$$P_{\xi, \beta(u)}(x) = 1 - \left(1 + \xi \frac{x}{\beta(u)}\right)^{-1/\xi}, \quad \beta(u) = \sigma + \xi(u - \mu), \quad (6)$$

pour tout  $x > 0$  tel que

$$1 + \xi \frac{x}{\beta(u)} > 0.$$

Lorsque  $\xi < 0$  la DPG est à support borné à droite et la borne est  $u - \beta/\xi$  ; lorsque  $\xi > 0$  la distribution n'a pas de borne supérieure. Si  $\xi = 0$ , la DPG est définie comme la limite de  $P_{\xi, \beta(u)}(x)$  quand  $\xi$  tend vers  $0^+$  (elle n'est pas bornée et décroît exponentiellement).

Pour un seuil  $u$  proche de la borne supérieure de la distribution, on peut utiliser la description paramétrique suivante de la queue de distribution :

$$F(x) = P\{X \leq x\} \stackrel{u \rightarrow x_0}{\longleftarrow} 1 - P\{X > u\} \left(1 + \xi \frac{x-u}{\beta}\right)^{-1/\xi}, \forall x / x - u > 0 \text{ et } 1 + \xi \frac{x-u}{\beta} > 0. \quad (7)$$

L'estimateur d'un quantile extrême  $Q_{1-\alpha} > u$  est obtenu en inversant l'expression de la fonction de distribution ci-dessus :

$$\widehat{Q}_{1-\alpha} = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left[ \left( \frac{\widehat{P}\{X > u\}}{1-\alpha} \right)^{\widehat{\xi}} - 1 \right],$$

où  $\widehat{\xi}$ ,  $\widehat{\beta}$  sont des estimées des paramètres du modèle DPG (voir par exemple [11]) et  $\widehat{P}\{X > u\}$  est estimé empiriquement par une méthode de Monte-Carlo.

Il est nécessaire de choisir  $u$  suffisamment grand pour se placer dans les conditions de validité asymptotique de (7) tout en gardant un nombre de données  $x_i > u$  suffisant pour obtenir une variance d'estimation pertinente (voir [11]).

## 5 Exemple

Soit  $\mathbf{X} = (R \cos \theta, R \sin \theta)^T$  où  $\theta \sim \mathcal{N}(0, \pi/10)$  et  $R$  suit une distribution t de Student à quatre degrés de liberté. Le tableau 1 compare les moyennes et écarts-type de  $P(\widehat{Q}_{0.995})$ , lorsque  $\widehat{Q}_{0.995}$  est estimé à partir de réalisations de  $\mathbf{X}$ , d'abord à l'aide de la densité de probabilité obtenue par classification SVM simple [1], puis par utilisation combinée de la classification SVM et de la théorie des valeurs extrêmes. On constate que la théorie des valeurs extrêmes fournit de bien meilleures estimées. La figure 2 représente 200 réalisations de  $\mathbf{X}$ , ainsi que le quantile empirique  $\widehat{Q}_{0.995}$  estimé à partir de ces réalisations et les intervalles de confiance, fournis par la théorie des valeurs extrêmes.

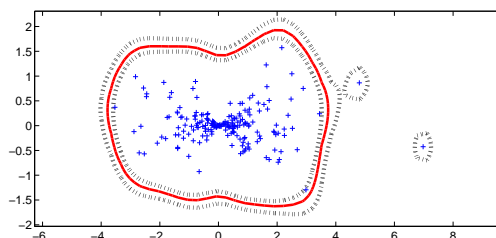


FIG. 2 – 200 échantillons du vecteur  $\mathbf{X}$ , le quantile à 0.995 estimé par théorie des valeurs extrêmes (représenté en trait continu) et les intervalles de confiance intérieurs et extérieurs (en traits pointillés). Le noyau est choisi gaussien avec  $\sigma = 0.75$

	$P(\mathbf{X} \in \hat{Q}_{0.995, \text{SVM}})$	$P(\mathbf{X} \in \hat{Q}_{0.995, \text{SVM+TVE}})$
$n = 100$	0.9338 (0.0205)	0.9903 (0.0117)
$n = 200$	0.961 (0.011)	0.9925 (0.0109)
$n = 400$	0.9765 (0.0063)	0.9933 (0.0042)

TAB. 1 – Comparaison, sur la base de 100 essais, des moyennes et écarts-type de  $P(\hat{Q}_{0.995})$ , lorsque  $\hat{Q}_{0.995}$  est estimé à l'aide de la seule SVM, puis par utilisation conjointe de SVM et TVE.

Cet exemple simple montre donc la pertinence de la méthode proposée. Les calculs de valeurs extrêmes multivariées sont en effet d'une importance cruciale en ingénierie, où les systèmes étudiés ont le plus souvent plusieurs sorties. Les vitesses de convergence des estimateurs obtenus en fonction de la dimension de  $\mathbb{X}$ , du type de décroissance de la queue, et du choix du noyau reproduisant sont des questions qu'il nous semble important d'étudier par la suite.

## Références

- [1] Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., and Williamson, R. (2001) Estimating the support of a high-dimensional distribution. *Neural Computation*, **13**, 1443–1471.
- [2] Einmahl, J. and Mason, D. (1992) Generalized quantile processes. *Annals of statistics*, **20**, 1062–1078.
- [3] Hartigan, J. (1987) Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.*, **82**, 267–270.
- [4] Nolan, D. (1991) The excess mass ellipsoid. *Journal of multivariate analysis*, **39**, 348–371.
- [5] Polonik, W. (1997) Minimum volume sets and generalized quantile processes. *Stochastic Processes and their Applications*, **69**, 1–24.
- [6] Tsybakov, A. (1997) On nonparametric estimation of density level sets. *Annals of Statistics*, **25**, 948–969.
- [7] Molchanov, I. (1990) Empirical estimation of distribution quantiles of random closed sets. *Theory Probab. Appl.*, **35**, 594–600.
- [8] Vert, R. and Vert, J.-P. (2005) Consistency and convergence rates of one-class svm and related algorithms. Tech. rep., Université Paris-Sud.
- [9] Aronszajn, N. (1950) Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, **68**, 337–404.
- [10] Schölkopf, B., Herbrich, R., and Smola, A. (2001) A generalized representer theorem. *Proceedings of the Annual Conference on Computational Learning Theory*, pp. 416–426.
- [11] Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.