

OUT-OF-DISTRIBUTION GENERALIZATION OF QUANTILE REGRESSION WITH HEAVY TAILED INPUTS : AN SVM APPROACH

Baptiste Leroux¹ & Clément Dombry² & Anne Sabourin³

¹ *Université Paris Cité, CNRS, MAP5, F-75006 Paris, France, baptiste.leroux@u-paris.fr*

² *Université Marie et Louis Pasteur, Laboratoire de mathématiques de Besançon*

³ *Université Paris Cité, Université Paris Saclay, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-75006, Paris, France*

Résumé La Théorie des Valeurs Extrêmes fournit des outils statistiques pour modéliser le comportement de variables aléatoires dans des conditions extrêmes, typiquement lorsque la norme d’une covariable franchit un seuil élevé. Dans de nombreux contextes (comme le changement climatique, avec des phénomènes de *covariate shift*) il est essentiel de construire des prédicteurs qui ont de bonnes propriétés d’extrapolation hors domaine. Des travaux récents montrent la pertinence de prédicteurs angulaires entraînés sur les observations les plus extrêmes sous des hypothèses de variation régulière, pour des tâches de classification binaire ou de régression par moindres carrés, dans le cadre limité de la minimisation du risque empirique. Nous considérons ici une procédure pénalisée, compatible avec la grande dimension : la régression quantile basée sur les *Support Vector Machines* (SVM). Nous établissons des garanties d’excès de risque non asymptotiques, et nous illustrons nos résultats avec un jeu de données multivariées concernant des mesures de débit fluvial dans le bassin du Danube.

Mots-clés Théorie des valeurs extrêmes, bornes d’erreur non-asymptotiques, régression quantile, Support Vector Machines

Abstract Extreme Value Theory (EVT) provides statistical tools to model the behavior of random variables under extreme conditions, typically when the norm of a covariate exceeds a high threshold. In many contexts (covariate-shifts, climate change), constructing predictors with good extrapolation properties is essential. Recent work has explored learning angular predictors from the most extreme observations under regular variation assumptions, in settings such as binary classification and least squares regression with empirical risk minimization. Here, we consider a penalized learning procedure, focusing on SVM-based quantile regression. We establish finite-sample learning guarantees and illustrate our results on a real life multivariate dataset involving river flow measurement on the Danube river network.

Keywords Extreme value theory, finite-sample error bounds, quantile regression, Support Vector Machines

1 Introduction

La régression quantile (Koenker and Bassett, 1978) est une stratégie de prédiction d’une variable cible réelle Y à partir d’une covariable X , consistant en l’estimation d’une fonction quantile conditionnelle de niveau $\tau \in (0, 1)$ de Y sachant X ,

$$q(x) = \inf\{t \in \mathbb{R} : \mathbb{P}(Y \leq t \mid X = x) \geq \tau\}.$$

Cette fonction quantile minimise le risque *pinball*, associé à une fonction de perte notée ℓ_τ . L’approche SVM (Steinwart and Christmann, 2008) de ce problème utilise une classe de prédicteurs formant un RKHS, et pénalise par la norme RKHS du prédicteur. L’estimateur h obtenu est le minimiseur du risque empirique pénalisé

$$\widehat{\mathcal{R}}^{(\lambda)}(h) = \frac{1}{n} \sum_{i=1}^n \ell_\tau(Y_i, f(X_i)) + \lambda \|h\|_{\mathcal{H}}^2.$$

Ici, cherchons à construire une fonction $\hat{q}(x)$ ayant de bonnes propriétés en extrapolation, lorsque $\|x\|$ prend une valeur inhabituellement grande. Il y a par construction peu de telles observations (X_i, Y_i) dans l’échantillon d’entraînement, et un algorithme basé sur la minimisation du risque empirique, ou sa version pénalisée pour les SVM, risque de ne pas fournir de prédiction fiable dans ces régions extrêmes, ou « hors-domaine » de la covariable X .

La généralisation hors-domaine basée sur la théorie des valeurs extrêmes et la variation régulière a été explorée par Cléménçon et al. (2025) dans le cadre minimaliste de la régression par moindres carrés. Des travaux connexes ont été menés dans le cadre de la classification binaire par Jalalzai et al. (2018). La régression quantile a quant à elle été abordée dans Buriticá and Engelke (2024) dans un cadre univarié et dans de Carvalho et al. (2022), dans un cadre d’estimation paramétrique adapté à l’étude des maxima plutôt que des excès de seuils.

Suivant la stratégie proposée dans les travaux pré-cités, étant donné un seuil t grand, nous considérons le risque pinball conditionné $\mathcal{R}_t(f) = \mathbb{E}[\ell_\tau(Y_i, f(X_i)) \mid \|X\| \geq t]$ et sa version empirique pénalisée dans un cadre SVM. Notre objectif n’est pas de minimiser un risque pinball classique $\mathcal{R}(f) = \mathcal{R}_0(f)$, prenant en compte les régions non extrêmes des covariables, mais une version asymptotique du risque pinball conditionnel,

$$\mathcal{R}_\infty(f) = \limsup_{t \rightarrow \infty} \mathcal{R}_t(f).$$

Les garanties théoriques existantes pour l’excès de risque asymptotique de prédicteurs multivariés concernent des cadres simplifiés, en supposant que la classe de prédicteurs est de dimension VC finie, et que la cible est bornée. Notre contribution principale est d’apporter des garanties théoriques pour une approche par minimisation d’un risque pénalisé qui permet de s’affranchir du cadre VC. Plus précisément, nous considérons un algorithme prédictif de type SVM. De plus, nous autorisons la variable cible à être non bornée, ce qui est rendu possible par le choix du cadre de la régression quantile plutôt que par moindres carrés. Cette extension est bienvenue dans le contexte de l’analyse des valeurs extrêmes, où les variables d’intérêt sont typiquement non bornées.

Le reste de ce document s'organise comme suit. La Section 2 présente la procédure de régression quantile SVM extrême proposée, ainsi que le cadre théorique dans lequel elle s'inscrit. Les garanties théoriques relatives à l'excès de risque de l'estimateur sont ensuite établies en Section 3. Enfin, la Section 4 illustre l'approche sur un cas d'étude.

2 Procédure proposée et cadre de la variation régulière

Les travaux de Cléménçon et al. (2025) et Jalalzai et al. (2018) établissent des garanties pour une stratégie n'utilisant que la partie angulaire des covariables les plus extrêmes. Nous adoptons une stratégie analogue, adaptée au cadre de la régression quantile, et considérons une version modifiée de l'algorithme quantile SVM de Takeuchi et al. (2006), menant au prédicteur

$$\hat{h}_\lambda \in \operatorname{argmin}_{h \in \mathcal{H}} \frac{1}{k} \sum_{i=1}^k \ell_\tau(Y_{(i)}, h \circ \theta(X_{(i)})) + \lambda \|h\|_{\mathcal{H}}^2$$

où $\mathcal{H}$ est un RKHS, $\|X_{(1)}\| \geq \|X_{(2)}\| \geq \dots \geq \|X_{(n)}\|$ sont les observations triées par ordre décroissant de leur norme, et $\theta(x) = x/\|x\|$. La fonction quantile estimée est alors $\hat{q} = \hat{h} \circ \theta$. Nous justifions maintenant la forme de l'estimateur proposé, et notamment sa nature angulaire. Comme dans Cléménçon et al. (2025); Jalalzai et al. (2018) nous faisons une hypothèse de *variation régulière par rapport à la covariable*.

Hypothèse 1.

$$(X/t, Y \mid \|X\| \geq t) \xrightarrow[t \rightarrow \infty]{\mathcal{L}} (X_\infty, Y_\infty).$$

Cette hypothèse permet de définir une fonction *quantile extrême*, notée q_{P_∞} , définie comme étant la fonction quantile associée au couple (X_∞, Y_∞) . Nous faisons une deuxième hypothèse de stabilité et de régularité sur q_{P_∞} .

Hypothèse 2. La fonction q_{P_∞} est continue et

$$\mathbb{E}[|q(X) - q_{P_\infty}(X)| \mid \|X\| \geq t] \xrightarrow[t \rightarrow \infty]{} 0.$$

Ces deux hypothèses justifient la nature angulaire du prédicteur cherché, et donc la forme du problème de minimisation défini précédemment, via le théorème suivant.

Théorème 1. *Lorsque les Hypothèses 1 et 2 sont vérifiées, on a :*

1. *La fonction quantile dans la limite des covariables extrêmes, q_{P_∞} , est angulaire, c'est-à-dire, pour tout $x \in \mathbb{R}^d \setminus \{0\}$, $q_{P_\infty}(x) = q_{P_\infty}(x/\|x\|)$.*
2. *La fonction q_{P_∞} est un minimiseur du risque asymptotique $\mathcal{R}_\infty$ parmi toutes les fonctions mesurables.*

L'exemple suivant montre que les Hypothèses 1 et 2 sont vérifiées dans des cadres standards en analyse des valeurs extrêmes.

Exemple 1. On considère un vecteur $(X, Z) \in \mathbb{R}^d \times \mathbb{R}$, avec $\mathbb{P}(X = 0) = 0$. On suppose que le vecteur (X, Z) a une densité π strictement positive et continue sur $\mathbb{R}^{d+1} \setminus \{0\}$. On suppose qu'il existe une fonction b à variation régulière et une densité π_∞ strictement positive, continue, telle que

$$\sup_{\|(x,z)\|>1} |t^{d+1}b(t)\pi(tx, tz) - \pi_\infty(x, z)| \xrightarrow[t \rightarrow \infty]{} 0.$$

Pour $x \in \mathbb{R}^d$, on définit $\pi_\infty(x) = \int_{\mathbb{R}} \pi_\infty(x, z) dz$ la densité marginale de X , et on suppose que cette densité est bornée inférieurement sur la sphère unité. Alors, la paire $(X, Z/\|X\|)$ vérifie les Hypothèses 1 et 2.

3 Borne sur l'excès de risque asymptotique

Le théorème qui suit fournit une décomposition de l'excès de risque asymptotique pour l'estimateur proposé. Cet excès de risque se décompose comme la somme d'un biais associé au choix de k pour la régression, d'une erreur d'approximation propre au choix du RKHS $\mathcal{H}$, et d'un terme de déviation.

Théorème 2. Soit $t_{n,k}$ le quantile $1 - \frac{k}{n}$ de $\|X\|$. Supposons que les Hypothèses 1 et 2 sont satisfaites, et que le noyau K associé au RKHS $\mathcal{H}$ est continu et borné par 1, i.e. $\sup_{x \in \mathbb{R}^d} K(x, x) = 1$. Alors,

$$\mathcal{R}_\infty(\widehat{h}_\lambda \circ \theta) + \lambda \|\widehat{h}_\lambda\|_{\mathcal{H}}^2 - \mathcal{R}_\infty^* \leq B(n, k, \lambda) + \mathcal{A}_2^\infty(\lambda) + D(n, k, \lambda)$$

où $B(n, k, \lambda)$ est un terme de biais dû au choix de k ,

$$B(n, k, \lambda) = 2 \sup_{\|h\|_{\mathcal{H}} \leq \frac{\max(\tau, 1-\tau)}{\lambda}} |\mathcal{R}_{t_{n,k}}(h \circ \theta) - \mathcal{R}_\infty(h \circ \theta)|,$$

le terme $\mathcal{A}_2^\infty(\lambda)$ est l'erreur d'approximation asymptotique du RKHS $\mathcal{H}$, définie par

$$\mathcal{A}_2^\infty(\lambda) = \inf_{h \in \mathcal{H}} \left(\mathcal{R}_\infty(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2 \right) - \mathcal{R}_\infty^*$$

et $D(n, k, \lambda)$ est un terme de déviation, défini par

$$D(n, k, \lambda) = \mathcal{R}_{t_{n,k}}^{(\lambda)}(\widehat{h}_\lambda \circ \theta) - \mathcal{R}_{t_{n,k}}^{(\lambda)}(h_{\lambda, t_{n,k}}^* \circ \theta),$$

où $h_{\lambda, t_{n,k}}^*$ est l'équivalent non-empirique de $\widehat{h}_\lambda$, défini par

$$h_{\lambda, t_{n,k}}^* \in \operatorname{argmin}_{h \in \mathcal{H}} \mathcal{R}_{t_{n,k}}(h \circ \theta) + \lambda \|h\|_{\mathcal{H}}^2.$$

Le choix du paramètre k met en jeu un compromis biais-variance, typique en analyse des valeurs extrêmes. Un k élevé réduit le terme de variance et le terme d'erreur d'approximation, mais augmente le biais. Le contrôle du terme $B(n, k, \lambda)$ est assuré par le théorème suivant.

Théorème 3. *Si k est choisi en fonction de n , $k = k(n)$, de sorte que $k(n)/n \rightarrow 0$, alors il existe une suite $\lambda_n \rightarrow 0$, dépendante du choix de $k(n)$, telle que*

$$B(n, k, \lambda_n) \xrightarrow[n \rightarrow \infty]{} 0.$$

Le théorème suivant fournit une borne en grande probabilité sur le terme $D(n, k, \lambda)$, ne dépendant que du nombre d'observations extrêmes k conservées et du paramètre λ .

Théorème 4. *Sous les mêmes hypothèses que précédemment, il existe une constante universelle C telle que, avec probabilité au moins $1 - \delta$,*

$$D(n, k, \lambda) \leq \frac{C \log(1/\delta)}{\lambda k}.$$

Enfin, sous des conditions de régularité sur la fonction q_{P_∞} (Eberts and Steinwart, 2012; Steinwart and Christmann, 2011), il est légitime de supposer que l'erreur d'approximation asymptotique vérifie $\mathcal{A}_2^\infty(\lambda) \leq c\lambda^\beta$ avec $c > 0, \beta > 0$. Ainsi, on obtient un contrôle sur le terme $\mathcal{A}_2^\infty(\lambda) + D(n, k, \lambda)$ en grande probabilité, de l'ordre de $k^{-\frac{\beta}{\beta+1}}$.

Remarque 1. Les contrôles sur le terme d'erreur $D(n, k, \lambda) + \mathcal{A}_2^\infty(\lambda)$ ainsi que sur le terme de biais $B(k, n, \lambda)$ assurent la consistance de la procédure proposée dans le cas où $k = k(n)$ avec $k(n) \rightarrow \infty$ et $k(n)/n \rightarrow 0$, pour un bon choix de $\lambda = \lambda(n)$.

4 Etude du débit fluvial dans le bassin du Danube

Le jeu de données étudié concerne 31 stations de mesures quotidiennes de débit dans le bassin du Danube. L'objectif est de prédire un quantile du débit en une station donnée, conditionnellement aux débits observés aux autres stations. Une attention particulière est portée à la fiabilité en situation de crise, lorsque au moins une station présente des valeurs extrêmes, typiquement lors d'inondations.

Les observations sont regroupées dans un vecteur $\tilde{X} \in \mathbb{R}^{31}$. La première étape, consiste en une standardisation pour que les lois marginales après transformation suivent une loi de Pareto unitaire. On note X_j le débit quotidien de la station j et F_j sa fonction de répartition. Le vecteur standardisé V sur lequel on travaillera est composé des observations $V_j = \left(\frac{1}{1-F_j(X_{ij})} \right)_{i \in \{1, \dots, n\}}$, et on note $V = (X, Z)$ où Z désigne la composante associée à la station cible et X celle correspondant aux autres stations. En posant $Y = Z/\|X\|$, comme dans l'Exemple 1, la paire (X, Y) vérifie les Hypothèses 1 et 2. La procédure précédemment présentée permet alors d'estimer les quantiles conditionnel de Y sachant $X = x$ par $\hat{q}(x)$, et d'en déduire les quantiles de Z sachant $X = x$ sous la forme $\|X\|\hat{q}(x)$.

En pratique, la queue de distribution de X_j au dessus d'un seuil u_j , est modélisée par une loi de Pareto généralisée, de sorte que, si ζ_j est la probabilité d'excès $\mathbb{P}(X_j > u_j)$, on a pour $x > u_j$, $F_j(x) = 1 - \zeta_j(1 + \xi_j \frac{x-u_j}{\sigma_j})^{-1/\xi_j}$. L'estimation des paramètres ξ_j, σ_j par maximum de vraisemblance, classiquement, permet d'obtenir une estimation de la distribution $\hat{F}_j$ au dessus du seuil u_j . En dessous de ce seuil u_j , la distribution F_j est estimée via

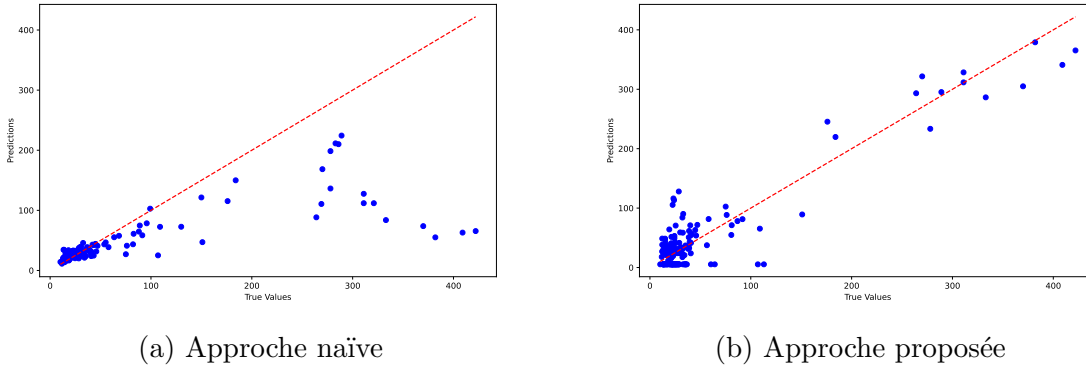


FIGURE 1 – Vraies valeurs de Z contre médiane prédite, sur les observations les plus extrêmes des données test, avec les deux approches susmentionnées.

$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_{i,j} \leq x\}$. Nous comparons notre proposition à une approche naïve, consistant à normaliser chaque composante de X avec sa moyenne et son écart type empirique, puis à apprendre des observations les plus extrêmes sur ce vecteur normalisé. La comparaison des deux approches, illustrée en Figure 1, montre que l’approche naïve échoue à fournir des prédictions fiables lorsque les observations sont grandes.

Références

- Buriticá, G. and Engelke, S. (2024). Progression : an extrapolation principle for regression. *arXiv preprint arXiv :2410.23246*.
- Cléménçon, S., Huet, N., and Sabourin, A. (2025). On regression in extreme regions. *Electronic Journal of Statistics*, 19(2) :4784–4828.
- de Carvalho, M., Kumukova, A., and Dos Reis, G. (2022). Regression-type analysis for multivariate extreme values. *Extremes*, 25(4) :595–622.
- Eberts, M. and Steinwart, I. (2012). Optimal regression rates for svms using gaussian kernels. *Electronic Journal of Statistics*, 6 :2627–2668.
- Jalalzai, H., Cléménçon, S., and Sabourin, A. (2018). On binary classification in extreme regions. *Advances in Neural Information Processing Systems*, 31.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1) :33–50.
- Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.
- Steinwart, I. and Christmann, A. (2011). Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17(1).
- Takeuchi, I., Le, Q. V., Sears, T. D., and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7 :1231–1264.