Prédictions de quantiles et d'expectiles spatiaux pour des champs elliptiques

Antoine Usseglio-Carleve [1] & Véronique Maume-Deschamps [2] & Didier Rullière [3]

[1] *Université de Lyon, Université Lyon 1, Institut Camille Jordan ICJ UMR 5208 CNRS*
usseglio@math.univ-lyon1.fr
[2] *Université de Lyon, Université Lyon 1, Institut Camille Jordan ICJ UMR 5208 CNRS*
veronique.maume@univ-lyon1.fr
[3] *Université de Lyon, Université Lyon 1, Laboratoire SAF EA2429*
didier.rulliere@univ-lyon1.fr

Résumé. Dans ce travail, nous considérons un champ elliptique. Nous proposons de prédire les quantiles et expectiles des valeurs du champ en un point, sachant qu'il a été observé en d'autres points. Le cas de la moyenne (expectile avec $\alpha = \frac{1}{2}$) est assez répandu dans la littérature. En effet, Krige (1951) a proposé d'exprimer cette moyenne conditionnelle comme combinaison linéaire des covariables observées, ce que Matheron (1963) formalisera plus tard comme le krigeage. En s'inspirant de cette idée, nous donnons les meilleurs prédicteurs affines des quantiles et expectiles du champ conditionnel, sous des critères que nous préciserons (voir Koenker et Bassett, 1978 et Newey et Powell, 1987). Les propriétés de stabilité par la somme des distributions elliptiques nous permettent également de donner la distribution de ces prédicteurs. En comparant ces prédicteurs aux valeurs théoriques (obtenues soit explicitement dans certains cas, soit par simulation), nous observons que le modèle affine peut être très éloigné de la réalité, surtout pour des quantiles (ou expectiles) extrêmes, i.e pour lesquels α est proche de 0 ou 1. Pour celà, nous proposons de nouveaux prédicteurs dits "extrêmes" et démontrons qu'ils sont asymptotiquement équivalents aux quantiles et expectiles théoriques lorsque α tend vers 0 et 1. A travers des exemples numériques, il apparait évident que les prédicteurs de régression sont moins efficaces lorsque l'on quitte le cadre gaussien, ce qui justifie l'utilisation des prédicteurs extrêmes. Ce travail est détaillé dans Maume-Deschamps et al. (2016a) et Maume-Deschamps et al. (2016b). Nous présentons en plus une application sur des données de température.

Mots-clés. Distributions elliptiques, Régression quantile, Régression expectile, Quantiles extrêmes, Expectiles extrêmes, Prédiction spatiale, Krigeage.

Abstract. In this work, we consider an elliptical random field. We propose some spatial quantile and expectile predictions at one site given observations of the field at some other locations. The case of the mean (expectile with $\alpha = \frac{1}{2}$) is fairly widespread in the literature. Indeed, Krige (1951) proposed to express this conditional mean as a linear combination of observed covariates, what Matheron (1963) will later formalize as kriging. On the basis of this idea, we give the best affine predictors of quantiles and expectiles of the conditional field under criteria that we will specify (see Koenker and Basset, 1978 and Newey and Powell, 1987). The properties of stability by the sum of elliptical distributions also allow us to give the distribution of these predictors. By comparing these predictors to theoretical values (obtained either explicitly in some cases

or by simulation), we observe that the affine model is not suitable in general, and can even be very far from reality, especially for extreme quantiles (or expectiles), ie for which α is close to 0 or 1. This is why we propose new "extreme" predictors and prove that they are asymptotically equivalent to the theoretical quantiles and expectiles when α tends to 0 and 1. Through numerical examples, the study shows that regression predictors may perform poorly when one leaves the usual Gaussian random field setting, justifying the use of extreme predictors. This work is detailed in Maume-Deschamps et al. (2016a) and Maume-Deschamps et al. (2016b).

Keywords. Elliptical distributions, Quantile regression, Expectile regression, Extremal quantiles, Extremal expectiles, Spatial prediction, Kriging.

Description

Les distributions elliptiques, définies par un vecteur $\mu \in \mathbb{R}^d$, une matrice $\Sigma \in \mathbb{R}^{d\times d}$, et une variable aléatoire positive R appelée rayon, admettent la représentation stochastique suivante (voir Cambanis et al., 1981) :

$$X = \mu + R\Lambda U^{(d)} \tag{1}$$

où $\Lambda\Lambda^T = \Sigma$ et $U^{(d)}$ est le vecteur aléatoire uniformément distribué sur la sphère de dimension d, indépendant de R. De plus, X est dite consistante, ou Gaussian Scale Mixtures, si $R \stackrel{\mathcal{L}}{=} \frac{\sqrt{\chi_d^2}}{\epsilon}$, avec ϵ une variable aléatoire positive indépendante de χ_d^2 (Kano, 1994), la loi du χ^2 à d degrés de liberté. Avec cette propriété de consistance, nous pouvons définir les champs elliptiques comme suit : un champ $(Z(t))_{t\in\mathbb{R}^d}$ est dit elliptique de rayon R si toute combinaison linéaire des $Z(t), t \in \mathbb{R}^d$, est elliptique de rayon R. Considérons maintenant que ce champ est observé en N points $t_1, ..., t_N \in \mathbb{R}^d$. Notre but est alors de prédire les quantiles et expectiles de la valeur $Z(t)$ du champ au point $t \in \mathbb{R}^d$. Pour celà, nous avons besoin d'informations sur la distribution conditionnelle des elliptiques. La thèse de Frahm (2004) en offre un bon résumé. En effet, si X est un vecteur elliptique de dimension d composé de deux sous-vecteurs X_1 et X_2 de dimensions respectives d_1 et d_2, $X_2|\,(X_1 = x_1)$ est toujours elliptique de vecteur $\mu_{2|1}$ et de matrice $\Sigma_{2|1}$ simples à obtenir. Notons que dans le cas gaussien, ces éléments sont respectivement la moyenne et la variance du vecteur conditionnel. En revanche, le rayon n'est plus R, mais R^* tel que

$$R^* \stackrel{\mathcal{L}}{=} R\sqrt{1-\beta}|\left(R\sqrt{\beta}U^{(d)} = C_{11}^{-1}(x_1 - \mu_1)\right) \tag{2}$$

où C_{11} est la racine de Cholesky de Σ_{11}, et $\beta \sim Beta\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$. Dès lors, si $X_2 \in \mathbb{R}$, les quantiles et expectiles conditionnels de niveau α, notés respectivement $q_\alpha\left(X_2|X_1 = x_1\right)$ et $e_\alpha\left(X_2|X_1 = x_1\right)$, valent :

$$\begin{cases} q_\alpha\left(X_2|X_1 = x_1\right) = \mu_{2|1} + \sqrt{\Sigma_{2|1}}\Phi_{R^*}^{-1}(\alpha) \\ e_\alpha\left(X_2|X_1 = x_1\right) = \mu_{2|1} + \sqrt{\Sigma_{2|1}}\Psi_{R^*}^{-1}(\alpha) \end{cases} \tag{3}$$

où $\Phi_{R^*}^{-1}(\alpha)$ et $\Psi_{R^*}^{-1}(\alpha)$ sont respectivement les quantiles et expectiles de la loi elliptique centrée réduite de rayon R^*. Nous disposons donc de formules théoriques. Malheureusement, l'expression de R^* donnée en Equation (2) est difficile à exploiter, et ne permet pas d'exprimer explicitement $\Phi_{R^*}^{-1}(\alpha)$ et $\Psi_{R^*}^{-1}(\alpha)$. Nous pouvons évidemment les approximer par des simulations ou des méthodes numériques, mais celà requiert un certain temps de calcul que nous cherchons à éviter. Pour celà, nous proposons de tester un prédicteur défini comme combinaison affine des covariables $Z = (Z(t_1), ..., Z(t_N))$, issu des régressions quantile (Koenker et Basset, 1978) et expectile (Newey et Powell, 1987) :

$$\begin{cases} \hat{q}_\alpha\left(X_2|X_1 = x_1\right) = \beta^{*T} Z + \beta_0^* \\ \hat{e}_\alpha\left(X_2|X_1 = x_1\right) = \delta^{*T} Z + \delta_0^* \end{cases} \tag{4}$$

où (β^*, β_0^*) et (δ^*, δ_0^*) sont respectivement définis comme suit :

$$\begin{cases} \underset{\beta\in\mathbb{R}^N,\beta_0\in\mathbb{R}}{\arg\min}\ (1-\alpha)\mathbb{E}\left[|Z(t) - \beta^T Z - \beta_0|\mathbb{1}_{\{Z(t)-\beta^T Z-\beta_0<0\}}\right] \\ \qquad +\alpha\mathbb{E}\left[|Z(t) - \beta^T Z - \beta_0|\mathbb{1}_{\{Z(t)-\beta^T Z-\beta_0>0\}}\right] \\ \underset{\delta\in\mathbb{R}^N,\delta_0\in\mathbb{R}}{\arg\min}\ (1-\alpha)\mathbb{E}\left[\left(Z(t) - \beta^T Z - \beta_0\right)^2 \mathbb{1}_{\{Z(t)-\beta^T Z-\beta_0<0\}}\right] \\ \qquad +\alpha\mathbb{E}\left[\left(Z(t) - \beta^T Z - \beta_0\right)^2 \mathbb{1}_{\{Z(t)-\beta^T Z-\beta_0>0\}}\right] \end{cases} \tag{5}$$

Dans le cadre des champs elliptiques, on est capable de résoudre le problème 5, ce qui mène aux prédicteurs de régression suivants (voir Maume-Deschamps et al., 2016a) et Maume-Deschamps et al., 2016b) :

$$\begin{cases} \hat{q}_\alpha\left(X_2|X_1 = x_1\right) = \mu_{2|1} + \sqrt{\Sigma_{2|1}}\Phi_R^{-1}(\alpha) \\ \hat{e}_\alpha\left(X_2|X_1 = x_1\right) = \mu_{2|1} + \sqrt{\Sigma_{2|1}}\Psi_R^{-1}(\alpha) \end{cases} \tag{6}$$

Ces prédicteurs utilisent donc les quantiles et expectiles de l'elliptique centrée réduite de rayon R, que l'on suppose connu. De plus, grâce à la stabilité par la somme des distributions elliptiques, les distributions de $\hat{q}_\alpha$ et $\hat{e}_\alpha$ sont obtenues facilement. Nous donnons des exemples de ces prédicteurs dans les cas gaussiens, Student, mélange gaussien unimodal, Laplace ou encore Slash. La comparaison de ces prédicteurs avec les valeurs théoriques semble montrer que dans le cas général, ces prédicteurs sont biaisés, particulièrement pour des valeurs de α extrêmes, i.e proches de 0 ou 1. Pour remédier à ce problème, nous proposons des nouveaux prédicteurs qui découlent de certaines hypothèses sur nos fonctions Φ_R et Ψ_R. En effet, on suppose qu'il existe $0 < \ell_1, \ell_2 < +\infty$ et $\gamma_1, \gamma_2 \in \mathbb{R}$ tels que :

$$\begin{cases} \lim\limits_{x\to+\infty} \frac{1-\Phi_{R^*}(x)}{1-\Phi_R(x^{\gamma_1})} = \ell_1 \\ \lim\limits_{x\to+\infty} \frac{1-\Psi_{R^*}(x)}{1-\Psi_R(x^{\gamma_2})} = \ell_2 \end{cases} \tag{7}$$

En ajoutant une hypothèse et des propriétés issues du papier de Djurcic et Torgasev (2001), on définit les prédicteurs suivants :

$$\begin{cases} \hat{\hat{q}}_{\alpha\uparrow}(X_2|X_1 = x_1) = \mu_{2|1} + \sqrt{\Sigma_{2|1}}\left[\Phi_R^{-1}\left(1 - \frac{1}{\frac{\ell_1}{1-\alpha}+2(1-\ell_1)}\right)\right]^{\frac{1}{\gamma_1}} \\ \hat{\hat{q}}_{\alpha\downarrow}(X_2|X_1 = x_1) = \mu_{2|1} - \sqrt{\Sigma_{2|1}}\left[\Phi_R^{-1}\left(1 - \frac{1}{\frac{\ell_1}{\alpha}+2(1-\ell_1)}\right)\right]^{\frac{1}{\gamma_1}} \\ \hat{\hat{e}}_{\alpha\uparrow}(X_2|X_1 = x_1) = \mu_{2|1} + \sqrt{\Sigma_{2|1}}\left[\Psi_R^{-1}\left(\frac{(2\alpha-1)\ell_2+1-\alpha}{(2\alpha-1)\ell_2+2(1-\alpha)}\right)\right]^{\frac{1}{\gamma_2}} \\ \hat{\hat{e}}_{\alpha\downarrow}(X_2|X_1 = x_1) = \mu_{2|1} - \sqrt{\Sigma_{2|1}}\left[\Psi_R^{-1}\left(\frac{(1-2\alpha)\ell_2+\alpha}{(1-2\alpha)\ell_2+2\alpha}\right)\right]^{\frac{1}{\gamma_2}} \end{cases} \tag{8}$$

On démontre par la suite que $\hat{\hat{q}}_{\alpha\uparrow}(X_2|X_1 = x_1)$ et $\hat{\hat{e}}_{\alpha\uparrow}(X_2|X_1 = x_1)$ sont asymptotiquement équivalents aux valeurs théoriques $q_\alpha(X_2|X_1 = x_1)$ et $e_\alpha(X_2|X_1 = x_1)$ lorsque α tend vers 1. De même pour $\hat{\hat{q}}_{\alpha\downarrow}(X_2|X_1 = x_1)$ et $\hat{\hat{e}}_{\alpha\downarrow}(X_2|X_1 = x_1)$ quand α tend vers 0.
Comme pour les prédicteurs de régression, nous testons nos prédicteurs extrêmes sur plusieurs exemples, et vérifions numériquement leur efficacité à l'aide d'outils tels que les $QQ-$plots ou $EE-$plots.

Bibliographie

[1] Cambanis, S., Huang, S. et Simons, G. (1981), On the theory of elliptically contoured distributions, *Journal of Multivariate Analysis*, 11, 368-385.

[2] Djurcic, D. et Torgasev, A. (2001), Strong Asymptotic Equivalence and Inversion of Functions in the Class Kc, *Journal of Mathematical Analysis and Applications*, 255, 383-390.

[3] Frahm, G. (2004), Generalized Elliptical Distributions: Theory and Applications, PhD thesis, Universitat zu Koln.

[4] Kano, Y. (1994), Consistency Property of Elliptical Probability Density Functions, *Journal of Multivariate Analysis*, 51, 139-147.

[5] Koenker, R. et Bassett, G. J. (1978), Regression Quantiles, *Econometrica*, 46(1), 33-50.

[6] Krige, D. (1951), A statistical approach to some basic mine valuation problems on the witwatersrand, *Journal of the Chemical, Metallurgical and Mining Society*, 52, 119-139.

[7] Matheron, G. (1963), *Traité de géostatistique appliquée*, Bureau de recherches géologiques et minières (France).

[8] Maume-Deschamps, V., Rullière, D. et Usseglio-Carleve, A. (2016a), Spatial Expectile Prediction for Elliptical Random Fields, preprint.

[9] Maume-Deschamps, V., Rullière, D. et Usseglio-Carleve, A. (2016b), Quantile Prediction for Elliptical Random Fields, à paraître dans *Journal of Multivariate Analysis*.

[10] Newey, W. et Powell, J. (1987), Asymmetric Least Squares Estimation and Testing, *Econometrica*, 55, 819-847.