Estimation non paramétrique des valeurs quantiles d'une série temporelle

Benoît Patra

Estimation non paramétrique des valeurs quantiles d'une série temporelle.

Benoît Patra

LSTA 175 rue du Chevaleret 75013 Paris, Lokad 9 rue Dareau 75014 Paris

Abstract

Time series forecasting applies to a large variety of problems. In order to forecast future values of a time series, it is frequently more robust to use an estimator based on the median or, more generally, based on a quantile. In this talk, we develop strategies for non-parametric sequential quantile forecasting. We prove the convergence of our strategies under weak assumptions when considering an expert-aggregation strategy relying on Nearest Neighbors experts. To conclude, those strategies are empirically evaluated against real world data - a call center call volume data set.
Keywords = Quantiles, Process statistics - time series.

Résumé

La prévision des séries temporelles couvre un vaste domaine d'applications statistiques. La prévision des valeurs ultérieures d'une série peut être effectuée par des estimateurs reposant sur la médiane et, plus généralement sur les quantiles. Dans cet exposé nous présentons des méthodes non-paramétriques de stratégies séquentielles de prévision quantile. Nous prouvons en particulier, sous des hypothèses faibles, la convergence d'une stratégie quantile reposant sur l' agrégation « d'experts ». Dans le cas présent, il s'agit d' estimateurs de type proches voisins. Nous illustrons nos résultats par des expériences menées sur des données réelles provenant de *call centers* (Centres d'appels téléphoniques).
Mots-clés = Quantiles, Statistique des processus - séries temporelles.

1 Introduction.

Une grande partie des données du monde réel se présente sous forme de séries temporelles. C'est-à-dire, une suite d'observations répétées d'une même variable à des temps différents. Les secteurs d'activités concernés par de tels types de données sont divers (banque, biotechnologies, grande distribution, énergie, ...). Afin de mieux comprendre le phénomène sous jacent ayant généré les observations, on utilise fréquemment une modélisation aléatoire. Ainsi, une série temporelle est bien souvent modélisée par un processus aléatoire à temps discret $(Y_n)_{n=-\infty}^{+\infty}$. Il s'agit d'une suite de variables aléatoires

que nous supposerons dans cet exposé dépendantes. Plusieurs objectifs peuvent être recherchés, le plus courant étant de prévoir les valeurs futures de la série. Dans ce cas, l'espérance conditionnelle $\mathbb{E}\left(Y_n|Y_1^{n-1}\right)$ est un prédicteur naturel où Y_1^{n-1} désigne la suite des variables du passé $(Y_1, \ldots, Y_{n-2}, Y_{n-1})$. La médiane conditionnelle est également un prédicteur pertinent du fait de sa robustesse aux observations aberrantes. Plus généralement, on peut chercher à prévoir la valeur du quantile conditionnel, à un seuil $\tau \in (0,1)$ fixé. Il s'agit du quantile d'ordre τ de la loi de la variable Y_n sachant Y_1^{n-1}, noté $q_\tau(Y_1^{n-1})$. Ceci permet d'estimer la borne supérieure avec probabilité τ de la valeur future de la série. Le quantile conditionnel d'ordre τ est défini par la formule suivante :

$$q_\tau(Y_1^{n-1}) = \inf\left\{q \in \mathbb{R} \text{ tels que } F_{Y_n|Y_1^{n-1}}(q) \geq \tau\right\},$$

où $F_{Y_n|Y_1^{n-1}}$ désigne la fonction de répartition conditionelle, c'est-à-dire,

$$F_{Y_n|Y_1^{n-1}}(x) = \mathbb{P}(Y_n \leq x|Y_1^{n-1}).$$

Notons $\rho_\tau : \mathbb{R} \to \mathbb{R}$ la fonction définie par $\rho_\tau(y) = y\left(\tau - \mathbb{I}_{\{y\leq 0\}}\right)$. Si Y_n est intégrable, le quantile $q_\tau(Y_1^{n-1})$ est solution du problème de minimisation,

$$\underset{q(.) \text{ mesurable}}{\operatorname{argmin}} \; \mathbb{E}_{Y_n|Y_1^{n-1}}\left(\rho_\tau\left(Y_n - q(Y_1^{n-1})\right)\right).$$

L'hypothèse (H) sur le processus assure l'unicité de ce minimum :
(H) : Presque sûrement, $F_{Y_0|Y_{-\infty}^{-1}}$ est continue strictement croissante.

2 Stratégies quantiles.

Nous supposons que la série temporelle $(y_1, \ldots, y_n, \ldots, y_N)$ provienne de la réalisation d'un processus $Y = (Y_n)_{n=-\infty}^{+\infty}$ stationnaire, ergodique et vérifiant l' hypothèse (H).
Une stratégie séquentielle de prévision quantile, que nous abrégerons en stratégie quantile, est une suite $g = (g_n)_{n=1}^{\infty}$, où, $g_n : \mathbb{R}^{n-1} \to \mathbb{R}$. Afin de mesurer les performances d'une telle stratégie ont utilise la fonction de perte cumulée et normalisée sur Y_1^n, définie par :

$$L_n(g) = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau\left(Y_t - g_t\left(Y_1^{t-1}\right)\right).$$

La définition suivante introduit la notion de convergence pour une stratégie séquentielle de prévision quantile.

Définition. *Une stratégie quantile est convergente par rapport à une classe $\mathcal{C}$ de processus ergodiques et stationnaires $(Y_n)_{-\infty}^{\infty}$ vérifiant l'hypothèse (H) si, pour chaque processus de la classe,*

$$\lim_{n\to\infty} L_n(g) = L^* \quad p.s.,$$

où,

$$L^* = \mathbb{E}\left(\min_{q(.)} \mathbb{E}_{Y_0|Y_{-\infty}^{-1}}\left(\rho_\tau\left(Y_0 - q(Y_{-\infty}^{-1})\right)\right)\right).$$

3 Une stratégie quantile convergente.

Nous définissons dans cette partie une stratégie quantile et prouvons sa convergence. L'estimateur d'indice k, ℓ, appelé expert dans ce context, au temps n, s'écrit :

$$h_n^{(k,\ell)}\left(y_1^{n-1}\right) = \underset{q\in\mathbb{R}}{\operatorname{argmin}} \sum_{t\in J_n^{(k,\ell)}} \rho_\tau(y_t - q).$$

La somme est effectuée sur l'ensemble des indices des plus proches voisins (PPV) :

$$J_n^{(k,\ell)} = \left\{k < t < n \text{ tels que } y_{t-k}^{t-1} \text{ est parmi les } \bar{\ell} \text{ PPV de } y_{n-k}^{n-1} \text{ parmi } y_1^k, \ldots, y_{n-k-1}^{n-2}\right\}.$$

Un tel expert effectue une recherche des $\bar{\ell}$ plus proches voisins du dernier segment de longueur k présent dans la série. Il retourne la valeur du quantile empirique calculé sur ces proches voisins. Notons que $\bar{\ell}$ est une simple fonction de ℓ et qu'une troncature sera nécessaire sur chaque expert afin de prouver la convergence de la méthode. L'estimateur final agrégé prend la forme de la combinaison convexe d'experts suivante :

$$g_n(y_1^{n-1}) = \sum_{k,\ell=1}^{\infty} p_{k,\ell,n} h_n^{(k,\ell)}(y_1^{n-1}),$$

où les poids $p_{k,\ell,n}$ d'un expert d'indice k, ℓ au temps n dépendent des perfomances passées de cet expert. Plus précisement, nous choisissons $b_{k,\ell} \geq 0$ vérifiant $\sum_{k,\ell} b_{k,\ell} = 1$, et posons :

$$w_{k,\ell,n} = b_{k,\ell} e^{-\eta_n (n-1) L_{n-1}(h^{(k,\ell)})}$$
$$p_{k,\ell,n} = \frac{w_{k,\ell,n}}{\sum_{i,j\geq 1}^{\infty} w_{i,j,n}}.$$

Le théorème suivant présente la convergence de cette stratégie de type proches voisins.

Théorème. *Soit $\mathcal{C}$ la classe des processus stationnaires ergodiques $(Y_n)_{n=-\infty}^{\infty}$ satisfaisant l'hypothèse (H) et $\mathbb{E}(Y_0^2) < +\infty$. Supposons que pour chaque entier $k \geq 1$ et chaque vecteur $\boldsymbol{s} \in \mathbb{R}^k$, la variable aléatoire $\left\|Y_1^k - \boldsymbol{s}\right\|$ a une fonction de distribution continue. Pour un choix de η_n tel que $\eta_n \to 0$ et $n\eta_n \to \infty$ lorsque $n \to \infty$,*
la stratégie de quantile définie précédemment est convergente par rapport à la classe $\mathcal{C}$.

La preuve de ce théorème s'appuie sur les travaux d'agrégation non paramétrique de Biau, Bleakley, Györfi et Ottucsák [1] ainsi que ceux de Györfi et Shäfer [3] dans le cadre d'une régression quadratique. Des éléments de preuve proviennent également de l'article [2] de Györfi, Udina et Walk, visant à l'optimisation de porte-feuilles d'actions.

Enfin une étude empirique de la stratégie quantile exposée sera effectuée sur des données provenant de *call centers* (centres d'appel). L'étude portera sur le volume d'appel entrant qui peut être décrit par une série temporelle. Cette étude comportera d'une part une mesure des performances de prédiction, d'autre part nous construirons des régions de confiance pour la série.

Bibliographie

[1] Biau Gérard, Bleakley Kevin et Györfi Lázló et Ottucsák György (2008). Nonparametric sequential prediction of time series, *Journal of Nonparametric Statistics*, London.
[2] Györfi Lázló, Udina Frederic et Walk Harro (2008). Nonparametric nearest neighbor based empirical portfolio selection strategies, *Statistics and Decisions*, Munich.
[3] Györfi Lázló et Shäfer Dustin (2003). Non parametric prediction. *Advances in Learning Theory : Methods, Models and Applications*