Consistance d'un estimateur de minimum de variance étendue

Consistency of a least extended variance estimator

Joseph Rynkiewicz [a]

[a] *SAMOS/MATISSE, Université de ParisI, 72 rue Regnault 75013 Paris, France*

Abstract

We consider a generalization of the criterion minimized by the K-means algorithm, where a neighborhood structure appears in the calculus of the variance. Such tool is used, for example with Kohonen maps, to measure the quality of the quantification preserving the neighborhood relationships. If we assume that the parameter vector is in a compact euclidean space and all it components are separated by a minimal distance, we show the strong consistency of the set of parameters almost realizing the minimum of the empirical extended variance.

Résumé

On considère une généralisation du critère minimisé par l'algorithme des K-moyennes [K-means], où une structure de voisinage apparait dans le calcul de la variance. Un tel outil est utilisé, par exemple avec des cartes de Kohonen, pour mesurer la qualité de la quantification respectant les structures de voisinage. Si on suppose que le vecteur paramètre est dans un compact d'un espace euclidien et que toutes ses composantes sont séparées par une distance minimale, on montre la consistance forte de l'ensemble des paramétres assez proches du minimum de variance étendue.

1. Introduction

Nous considérons une généralisation de la variance intra-classe qui est considérée comme le principal critère de mesure de qualité des cartes de Kohonen (cf Kohonen [4]), bien que l'algorithme de Kohonen ne minimise pas exactement ce critère (cf Cottrell et al. [1]). La variance étendue est la somme de la variance intra-classe et d'un terme qui dépend des classes voisines. Sa minimisation permet notamment d'obtenir une classification, qui respecte les relations de voisinage et qui donne lieu à des interprétations aisées, puisque la proximité des classes correspond à la proximité des données réelles représentées par ces classes.

Email address: `joseph.rynkiewicz@univ-paris1.fr` (Joseph Rynkiewicz).

Nous considérons dans toute la suite que les observations sont dans le compact $[0,1]^d$, qu'elles ont pour mesure de probabilité P qui admet une densité par rapport à la mesure de Lebesque bornée par une constante B.

Définition 1.1 *Pour $e \in \mathbb{N}^*, e \leq d$, soit un ensemble fini $I \subset \mathbb{Z}^e$ et Λ la fonction de voisinage définie de $I - I := \{i-j, i, j \in I\}$ dans $[0,1]$ telle que $\Lambda(k) = \Lambda(-k)$ et $\Lambda(0) = 1$.*

Définition 1.2 *Notons $\|.\|$ la norme euclidienne, soit*

$$D_I^\delta := \left\{ x := (x_i)_{i\in I} \in \left([0,1]^d\right)^I \text{ tels que } \|x_i - x_j\| \geq \delta, \text{ si } i \neq j \right\}$$

L'ensemble des centroïdes x_i séparés par une distance d'au moins δ.

Définition 1.3 *La tessellation de Voronoï $(C_i(x))_{i\in I}$ est définie par*

$$C_i(x) := \left\{ \omega \in [0,1]^d \text{ tels que } \|x_i - \omega\| < \|x_j - \omega\| \text{ si } j \neq i \right\}$$

de même, l'indice de la tesselation Voronoï pour une observation ω est définie par

$$C_x^{-1}(w) := \{ i \in I \text{ tels que } \|x_i - \omega\| < \|x_j - \omega\| \text{ si } j \neq i \}$$

En cas d'égalité, on assigne $\omega \in C_i(x)$ (resp. $i \in C_x^{-1}(\omega)$) grâce à l'ordre lexicographique.

Définition 1.4 *La variance étendue est*

$$V(x) := \frac{1}{2} \sum_{i,j\in I} \Lambda(i-j) \int_{C_i(x)} \|x_j - \omega\|^2 dP(\omega)$$

De même, lorsqu'il y a une nombre fini, n, d'observations, on définit la variance étendue empirique :

$$V_n(x) := \frac{1}{2n} \sum_{i\in I} \sum_{\omega \in C_i(x)} \left(\sum_{j\in I} \Lambda(i-j) \|x_j - \omega\|^2 \right)$$

Comme la variance étendue empirique n'est pas une fonction continue en x, il n'existe pas, en général, d'ensemble de centroïdes x_n réalisant le minimum de $V_n(x)$. La question que nous proposons d'étudier est donc la suivante : Si on considère les suites x_n telles que $V_n(x_n)$ soit suffisamment proche de son minimum, ces suites convergent-elles vers l'ensemble des centroïdes minimisant la variance théorique $V(x)$. Pour cela, nous procédons selon le même schéma de démonstration que Pollard [5] et nous commençons par montrer que les fonctions de variance étendue vérifient une loi uniforme des grands nombres.

2. Loi uniforme des grands nombres

Soit la famille de fonction

$$\mathcal{G} := \left\{ g_x(\omega) := \sum_{j\in I} \Lambda\left(C_x^{-1}(\omega) - j\right) \|x_j - \omega\|^2 \text{ pour } x \in D_I^\delta \right\}$$

Pour montrer la loi uniforme des grands nombres, il suffit de montrer que

$$\sup_{x\in D_I^\delta} \left| \int g_x(\omega) dP_n(\omega) - \int g_x(\omega) dP(\omega) \right| \xrightarrow{a.s} 0 \tag{1}$$

puisque, pour toute mesure de probabilité Q sur $[0,1]^d$:

$$\int g_x(\omega)dQ(\omega) = \int \sum_{j\in I} \Lambda\left(C_x^{-1}(w) - j\right) \|x_j - \omega\|^2 dQ(\omega) = \frac{1}{2}\sum_{i,j\in I} \Lambda(i-j) \int\limits_{C_i(x)} \|x_j - \omega\|^2 dQ(\omega)$$

d'après Gaenssler et Stute [3], une condition suffisante pour que l'équation (1) soit vérifiée est que : $\forall\varepsilon > 0, \forall x_0 \in D_I^\delta$ il existe un voisinage $S(x_0)$ de x_0 tel que

$$\int g_{x_0}(\omega)dP(\omega) - \varepsilon < \int \left(\inf_{x\in S(x_0)} g_x(\omega)\right) dP(\omega) \leq \int \left(\sup_{x\in S(x_0)} g_x(\omega)\right) dP(\omega) < \int g_{x_0}(\omega)dP(\omega) + \varepsilon$$

On peut d'abord prouver le résultat suivant, en utilisant une technique similaire à la preuve du lemme 11 de Fort et Pagès [2].

Lemme 2.1 *Soit $x \in D_I^\delta$ et λ la mesure de Lebesgue sur $[0,1]^d$. Notons E^c le complémentaire de l'ensemble E dans $[0,1]^d$ et $|I|$ le cardinal de l'ensemble I. Pour $0 < \alpha < \frac{\delta}{2}$, soit*
$U_i^\alpha(x) = \left\{\omega \in [0,1]^d / \exists y \in D_I^\delta, x_j = y_j \text{ si } j \neq i \text{ et } \|x_i - y_i\| < \alpha \text{ et } \omega \in C_i^c(y) \cap C_i(x)\right\}$
l'ensemble des w changeant de cellule de Voronoï lorsque le centroïde x_i se déplace d'une distance d'au plus α. Alors

$$sup_{x\in D_I^\delta} \lambda\left(U_i^\alpha(x)\right) < (|I|-1)\left(\frac{2\alpha}{\delta} + \alpha\right)\left(\sqrt{2}\right)^{d-1}$$

Considérons maintenant $x^0 \in D_I^\delta$ et $S(x^0)$ un voisinage de x^0 inclus dans une boule de rayon α, pour la distance euclidienne sur D_I^δ. Soit $W(x^0)$ l'ensemble des ω restant dans leur cellule de Voronoï lorsque on déplace x^0 vers n'importe quel $x \in S_{x_0}$. Pour tout $\omega \in W(x^0)$ on a

$$\begin{aligned}
\inf_{x\in S(x^0)} g_x(\omega) &\geq g_{x^0}(\omega) - \sum_{j\in I} \Lambda\left(C_{x^0}^{-1}(\omega) - j\right) \left(\|x_j^0 - \omega\|^2 - \inf_{x\in S(x^0)} \|x_j^0 - \omega\|^2\right) \\
&\geq g_{x_j^0}(\omega) - \sum_{j\in I} \left(\|x_j^0 - \omega\|^2 - \inf_{x\in S(x^0)} \|x_j^0 - \omega\|^2\right)
\end{aligned}$$

Pour tout $\omega \in [0,1]^d$, on a, pour α suffisamment petit, $\left(\|x_j^0 - \omega\|^2 - \inf_{x\in S(x^0)} \|x_j - \omega\|^2\right) < \frac{\varepsilon}{2B|I|}$ ainsi

$$\int\limits_{W(x^0)} \sum_{j\in I} \left(\|x_j^0 - \omega\|^2 - \inf_{x\in S(x^0)} \|x_j - \omega\|^2\right) dP(\omega) < \frac{\varepsilon}{2} \text{ et } \int\limits_{W(x^0)} \left(g_{x^0}(\omega) - \inf_{x\in S(x^0)} g_x(\omega)\right) < \frac{\varepsilon}{2}$$

Soit, maintenant $W(x^0)^c$, l'ensemble des ω changeant de cellule de Voronoï quand les centroïdes vont de x^0 vers un $x \in S_{x^0}$. Si $\alpha < \frac{\delta}{2|I|}$, alors en déplaçant séquentiellement les composantes x_i^0 de x^0 vers x_i de x, chaque configuration intermédaire reste dans $D_I^{\frac{\delta}{2}}$. Comme, pour tout $i \in I$, $\|x_i - \omega\|^2$ est borné par 1 sur $[0,1]^d$, le lemme 2.1, assure alors que

$$\int\limits_{W(x^0)^c} g_x(\omega)dP(\omega) < B|I|(|I|-1))\left(\frac{4\alpha}{\delta} + \alpha\right)\left(\sqrt{2}\right)^{d-1}$$

Finalement, si on choisit α suffisamment petit pour que $B|I|(|I|-1))\left(\frac{4\alpha}{\delta} + \alpha\right)\left(\sqrt{2}\right)^{d-1} < \frac{\varepsilon}{2}$, on obtient

$$\int\limits_{D_I^\delta} g_{x^0}(\omega)dP(\omega) - \varepsilon < \int\limits_{D_I^\delta} \left(\inf_{x\in S(x^0)} g_x(\omega)\right) dP(\omega)$$

Exactement de la même façon, pour α suffisamment petit, on obtient :

$$\int_{D_I^\delta} \left(\sup_{x \in S(x^0)} g_x(\omega) \right) dP(\omega) < \int_{D_I^\delta} g_{x^0}(\omega) dP(\omega) + \varepsilon$$

Ainsi, la condition suffisante pour la loi uniforme des grands nombres est vraie pour la variance étendue.

3. Consistance

Soit l'ensemble des "quasi-estimateurs" de minimum de variance étendue :

$$\bar{\chi}_n^\beta := \left\{ x \in D_I^\delta \text{ tels que } V_n(x) < \inf_{x \in D_I^\delta} V_n(x) + \frac{1}{\beta(n)} \right\}$$

avec $\beta(n)$ une fonction strictement positive tel que $\lim_{n\to\infty} \beta(n) = \infty$ Soit $\bar{\chi} = \arg\min_{x \in D_I^\delta} V(x)$ L'ensemble qui minimise la variance étendue théorique, comme la fonction $x \longmapsto V(x)$ est continue et non constante sur D_I^δ, pour tout voisinage $\mathcal{N}$ de $\bar{\chi}$, il existe $\eta(\mathcal{N}) > 0$ tel que

$$\forall x \in D_I^\delta \backslash \mathcal{N},\ V(x) > \min_{x \in D_I^\delta} V(x) + \eta(\mathcal{N})$$

Pour montrer la consistance forte, il suffit de montrer que pour tout voisinage $\mathcal{N}$ de $\bar{\chi}$ on a

$$\lim_{n\to\infty} \bar{\chi}_n^\beta \overset{p.s}{\subset} \mathcal{N} \Longleftrightarrow \lim_{n\to\infty} V\left(\bar{\chi}_n^\beta\right) - V(\bar{\chi}) \overset{p.s.}{\leq} \eta(\mathcal{N})$$

avec $V(E) - V(F) := \sup\{V(x) - V(y) \ \text{ pour } \ x \in E \ \text{ et } \ y \in F\}$.

Par définition $V_n\left(\bar{\chi}_n^\beta\right) \leq V_n(\bar{\chi}) + \frac{1}{\beta(n)}$, de plus la loi uniforme des grands nombres assure que $\lim_{n\to\infty} V_n(\bar{\chi}) - V(\bar{\chi}) \overset{p.s}{=} 0$, on obtient ainsi $\lim_{n\to\infty} V_n\left(\bar{\chi}_n^\beta\right) \overset{p.s.}{\leq} V(\bar{\chi}) + \frac{\eta(\mathcal{N})}{2}$ De même on aura $\lim_{n\to\infty} V\left(\bar{\chi}_n^\beta\right) - V_n\left(\bar{\chi}_n^\beta\right) \overset{p.s.}{=} 0$ et

$$\lim_{n\to\infty} V\left(\bar{\chi}_n^\beta\right) - \frac{\eta(\mathcal{N})}{2} \overset{p.s.}{<} \lim_{n\to\infty} V_n\left(\bar{\chi}_n^\beta\right) \overset{p.s.}{\leq} V(\bar{\chi}) + \frac{\eta(\mathcal{N})}{2}$$

finalement $\lim_{n\to\infty} V\left(\bar{\chi}_n^\beta\right) - V(\bar{\chi}) \overset{p.s.}{\leq} \eta(\mathcal{N})$ ce qui prouve la consistance forte du quasi-estimateur de minimum de variance étendue.

Preuve du lemme 2.1

Soit x et $y \in D_I^\delta$ vérifiant les hypothèses du lemme et $j \neq i \in I$, pour prouver cette inégalité, il suffit de borner la mesure des ω appartenant aux cellules $C_i(x)$ et $C_j(y)$ simultanément, puisque $(C_i(y))^c = \bigcup_{j \in I, j \neq i} C_j(y)$.

Notons $(z\,|t)$, le produit scalaire entre z et t, et $\overrightarrow{n}_x^{ij} := \frac{x_j - x_i}{\|x_j - x_i\|}$. Le vecteur paramètre $x + \gamma_1 \overrightarrow{n}_x^{ij}$ désignera le vecteur dont toutes les composantes sont égales à celles de x sauf la composante i qui sera égale à $x_i + \gamma_1 \overrightarrow{n}_x^{ij}$.

Comme $\|y_i - x_i\| < \alpha$, on aura $\left(y_i - x_i \,\middle|\, \overrightarrow{n}_x^{ij}\right) = \gamma_1$ avec $|\gamma_1| \leq \alpha < \frac{\delta}{2}$. Puisque la mesure de Lebesque (de $\mathbb{R}^{d-1}$) de toute section plane de $[0,1]^d$ est bornée par $\left(\sqrt{2}\right)^{d-1}$, la mesure de Lebesgue des ω changeant de cellule de Voronoï pour un mouvement, du centroïde x_i, de $\gamma_1 \overrightarrow{n}_x^{ij}$ est donc bornée par $\frac{|\gamma_1|}{2}\left(\sqrt{2}\right)^{d-1}$, d'où

$$\lambda\left(C_j\left(x+\gamma_1 \overrightarrow{n}_x^{ij}\right)\cap C_i(x)\right) < \alpha\left(\sqrt{2}\right)^{d-1}$$

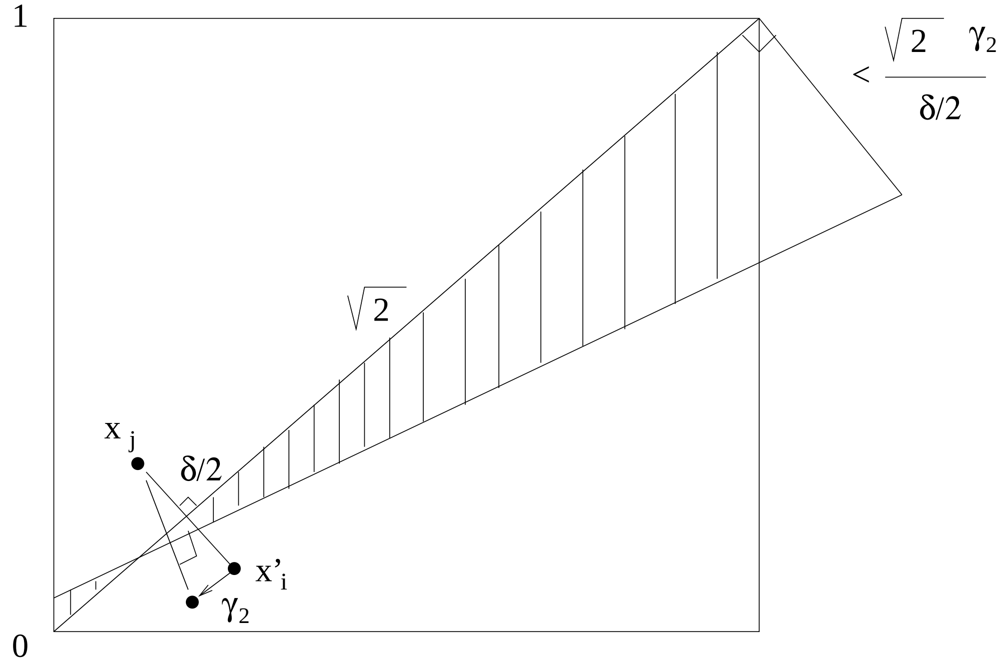
De plus, on peut remarquer que $x+\gamma_1 \overrightarrow{n}_x^{ij}$ appartient à $D_I^{\frac{\delta}{2}}$.

D'autre part, soit $y_i - x_i - \gamma_1 \overrightarrow{n}_x^{ij} := \gamma_2 \overrightarrow{\tau}_x^{ij}$, avec $\|\overrightarrow{\tau}_x^{ij}\| = 1$, la composante orthogonale à $\overrightarrow{n}_x^{ij}$ du déplacement de x_i vers y_i, c'est à dire telle que $\left(\overrightarrow{n}_x^{ij}\middle|\overrightarrow{\tau}_x^{ij}\right)=0$.

Comme l'illustre la figure (1), en dimension 2, pour tout $x' = x+\gamma_1 \overrightarrow{n}_x^{ij} \in D_I^{\frac{\delta}{2}}$, la mesure de Lebesgue des ω changeant de cellule de Voronoï pour un mouvement, du centroïde x'_i, de $\gamma_2 \overrightarrow{\tau}_x^{ij}$ est bornée par $\frac{2\alpha}{\delta}\left(\sqrt{2}\right)^{d-1}$, on aura ainsi

$$\lambda\left(C_j\left(x+\gamma_1 \overrightarrow{n}_x^{ij}+\gamma_2 \overrightarrow{\tau}_x^{ij}\right)\cap C_i(x)\right) < \alpha\left(\sqrt{2}\right)^{d-1} + \frac{2\alpha}{\delta}\left(\sqrt{2}\right)^{d-1}$$

FIG. 1. aire hachurée $< \frac{2\gamma_2}{\delta} < \frac{\sqrt{2}\times 2\alpha}{\delta}$



comme ces inégalités ne dépendent pas du choix de x, on obtient finalement

$$\sup_{x\in D_I^{\delta}} \lambda\left(C_j\left(x+\gamma_1 \overrightarrow{n}_x^{ij}+\gamma_2 \overrightarrow{\tau}_x^{ij}\right)\cap C_i(x)\right) < \left(\alpha+\frac{2\alpha}{\delta}\right)\left(\sqrt{2}\right)^{d-1}$$

d'où

$$\sup_{x\in D_I^{\delta}} \lambda\left(U_i^{\alpha}(x)\right) < \left(|I|-1\right)\left(\alpha+\frac{2\alpha}{\delta}\right)\left(\sqrt{2}\right)^{d-1}$$

■

Références

[1] Cottrell, M., Fort, J.C. and Pagès, G., Theoretical aspects of the SOM algorithm, Neurocomputing, volume 21, p. 119-138, 1998

[2] Fort, J.C and Pagès, G., On the A.S. convergence of the Kohonen algorithm with a general neighborhood function, Ann. Appl. Prob., volume 5 :4, p.1177-1216, 1995

[3] Gaenssler, P., Stute, W., Empirical processes : A survey of results for independent and identically distributed random variables, Ann. Prob., volume 7 :2 (1979), p.193-243

[4] Kohonen, T., Self-Organizing Maps, Springer Series in Information Sciences, Volume 30, Springer, 1995.

[5] Pollard, D., Strong consistency of k-mean clustering, Ann. Stat., volume 9 :1 (1981), p. 135-140