ALGORITHMES D'APPROXIMATION STOCHASTIQUE EN ANALYSE DES DONNEES

Jean-Marie Monnez

Institut Elie Cartan de Nancy

NANCY Université - CNRS - INRIA

On a à traiter actuellement des masses de données statistiques de plus en plus importantes en ce qui concerne le nombre d'individus ou le nombre de variables.

Les données peuvent arriver dans le temps et les modèles d'analyse évoluer dans le temps.

L'utilisation d'algorithmes d'approximation stochastique permet de réaliser séquentiellement des estimations, par exemple de paramètres de modèles de régression ou de facteurs d'analyse factorielle.

On présente dans cet exposé quelques algorithmes classiques adaptés à des problèmes d'analyse des données.

PLAN

1. Modèle linéaire
2. Algorithme des k-means
3. Approximation stochastique de vecteurs propres
4. ACP séquentielle
5. ACP d’un vecteur aléatoire
6. Méthode de gradient
7. Approximation stochastique des facteurs d’une ACP
8. Variantes
9. Analyse factorielle d’un vecteur aléatoire « multiple »
10. Solution de l’analyse canonique généralisée
11. Solution de l’analyse factorielle multiple
12. ACP partielle d’un vecteur aléatoire
13. ACG partielle d’un vecteur aléatoire
14. Extensions
15. ACP projetée et cas particuliers

1. MODELE LINEAIRE

Modèle linéaire

$(Y_1,\ldots,Y_n,\ldots)$ suite de variables aléatoires réelles observées :
pour $n \geq 1$, $Y_n = b_n' \theta + R_n$,
b_n vecteur connu au temps n de dimension p,
θ vecteur inconnu de dimension p,
R_n variable aléatoire réelle résiduelle telle que $E[R_n] = 0$.

Processus (X_n) d'approximation stochastique de θ (Albert et Gardner, 1967)

$$X_{n+1} = X_n - a_n b_n (b_n' X_n - Y_n).$$

Processus sous contraintes convexes

Dans certains modèles, θ appartient à un convexe fermé K de R^p, par exemple :

. régression sous contrainte de positivité des coefficients ;

. régression ridge ($\|\theta\|_2 \leq c$) ;

. régression LASSO ($\|\theta\|_1 \leq c$).

Soit Π l'opérateur de projection sur K.

$$X_{n+1} = \Pi(X_n - a_n b_n (b_n' X_n - Y_n)).$$

2. ALGORITHME DES k-MEANS DE MacQueen (1)

Soit un vecteur aléatoire réel Z de dim p défini sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

On cherche à partitionner Ω en r classes à partir des observations de Z.

<u>Critère de classification non supervisée</u>

Soit $I_{k\pi}$ l'indicatrice du $k^{ième}$ élément d'une partition π de Ω.

On cherche une partition $\pi = (\Omega^1, \ldots, \Omega^r)$ et un r-uplet de points $(\theta^1, \ldots, \theta^r)$ qui rendent minimale

$$E\left[\sum_{k=1}^{r} I_{k\pi} \left\| Z - x^k \right\|^2\right].$$

Si $\theta^1, \ldots, \theta^r$ sont connus, le minimum est atteint en prenant pour $k = 1, \ldots, q$,

$$\Omega^k = \{ \omega : ||Z(\omega) - \theta^k|| = min_j\ ||Z(\omega) - \theta^j|| \}$$

On cherche alors $\theta^1, \ldots, \theta^r$ qui rendent minimale la fonction

$$g\left(x^1, \ldots, x^r\right) = g(x) = E\left[\min_k \left\| Z - x^k \right\|^2\right] = E\left[\sum_{j=1}^{r} I_j(Z; x) \left\| Z - x^j \right\|^2\right]$$

$$\text{avec} \quad I_k(Z(\omega); x) = \begin{cases} 1 \text{ si } \left\| Z(\omega) - x^k \right\| \leq \left\| Z(\omega) - x^j \right\|, \ j = 1, \ldots, r \\ 0 \text{ sinon.} \end{cases}$$

2. ALGORITHME DES k-MEANS DE MacQueen (2)

$(\theta^1,\ldots,\theta^r)$ est une solution du système d'équations :

$$\frac{\partial g}{\partial x^k} = 2\,E\left[I_k(Z;x)\left(x^k - Z\right)\right] = 0,\ k = 1,\ldots,r.$$

Processus d'approximation stochastique $(X_n) = ((X_n^1, \ldots, X_n^r))$

$$X_{n+1}^k = X_n^k - a_n^k\ I_k(Z_n\ ;\ X_n)\ (X_n^k - Z_n),\ k = 1,\ \ldots,r.$$

a_n^k est une variable aléatoire réelle positive mesurable par rapport à la tribu du passé engendrée par $X_1, Z_1, \ldots, Z_{n-1}$.

Pour un choix particulier de a_n^k, on retrouve l'algorithme des *k*-means de MacQueen (1967) :

$$a_n^k = 1/\ (1 + J_k(n)\ ),\ J_k(n) = card\ \{\ j<n : I_k(Z_j\ ;\ X_j) = 1\ \}.$$

Le critère de classification peut être généralisé en remplaçant dans la définition du critère

$||Z - x^k||^2$ par une mesure de dissimilarité $D_k(Z\ ;\ x^k)$. La convergence presque sûre du processus d'approximation stochastique correspondant vers l'ensemble des points stationnaires de *g* est étudiée dans (Bouamaine, 1996).

3. APPROXIMATION STOCHASTIQUE DE VECTEURS PROPRES (1)

Algorithme de BENZECRI

Soit une matrice aléatoire *A* réelle carrée d'ordre *p*.

On suppose $B = E[A]$ *M*-symétrique (*M* connue), semi-définie positive de valeurs propres $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.

Soit (A_n) un échantillon i.i.d. de *A*.

Soit le processus d'approximation stochastique (X_n) (Benzécri, 1969) :

$$X_{n+1} = \left(I + \frac{1}{n^{\gamma}} A_n \right) X_n, \quad \frac{5}{6} < \gamma \leq 1.$$

(X_n) converge presque sûrement en direction vers un vecteur propre associé à la plus grande valeur propre λ_1.

Benzécri démontre la convergence presque sûre en direction d'un processus du même type vers un vecteur propre associé à la $k^{ième}$ plus grande valeur propre, $k = 2, 3, \ldots, p$, en faisant une orthogonalisation par rapport à la métrique *M*.

3. APPROXIMATION STOCHASTIQUE DE VECTEURS PROPRES (2)

Extension

En analyse factorielle, en général :

. la métrique M est inconnue

. A n'est pas directement observable.

Par exemple :

$A = M^{-1} (Z - E[Z])(Z - E[Z])'$,

$B = E[A] = M^{-1}$ *Covar* $[Z]$,

la métrique M dépend de la loi inconnue de Z.

Monnez (1994) et Bouamaine et Monnez (1997, 1998) établissent la convergence presque sûre en direction vers des vecteurs propres de B associés aux valeurs propres rangées par ordre décroissant de processus du type

$$Y_{n+1}^{i} = X_n^i + a_n \left(B_n - F_n(X_n^i) I \right) X_n^i, \quad i = 1, \dots, r,$$
$$\left(X_{n+1}^1, \dots, X_{n+1}^r \right) = orth_{M_n} \left(Y_{n+1}^1, \dots, Y_{n+1}^r \right).$$

. $E[B_n | T_n] \xrightarrow{p.s.} B$, T_n tribu du passé au temps n,

. F_n fonction aléatoire de R^p dans R,

. $M_n \xrightarrow{p.s.} M.$

3. APPROXIMATION STOCHASTIQUE DE VECTEURS PROPRES (3)

Ceci peut être utilisé pour estimer séquentiellement par exemple

. des facteurs principaux

. des axes principaux

de différentes analyses factorielles :

. ACP, AFCM, ACG, AFM, AC, AFD, AFC,….

Nous allons en présenter quelques exemples.

4. ACP SEQUENTIELLE (1)

On se place ici dans le cas où l'on dispose d'une suite de vecteurs de données $z_1,\ldots,z_n,\ldots$. arrivant séquentiellement dans le temps.

Modèle 1. On suppose que $z_1,\ldots,z_n,\ldots$ constituent un échantillon i.i.d. d'un vecteur aléatoire Z défini sur un espace probabilisé (Ω,A,P).

On peut définir une ACP de ce vecteur aléatoire qui représente l'ACP effectuée sur la population Ω.

Soit θ un résultat de l'ACP, par exemple une valeur propre, un facteur, un axe principal. Le problème est d'estimer θ.

Comme les données arrivent séquentiellement, on peut effectuer une estimation récursive de θ en utilisant un processus d'approximation stochastique :

x_1 est une estimation a priori de θ ou est choisi de façon aléatoire ;
disposant d'une estimation x_n de θ obtenue à partir des observations $z_1,\ldots,z_{n-1}$,
on introduit l'observation z_n
et on définit à partir de x_n et z_n une nouvelle estimation x_{n+1} de θ :

$$x_{n+1}=f_n(x_n,z_n).$$

4. ACP SEQUENTIELLE (2)

Modèle 2. On considère le cas où l'espérance mathématique de Z_n varie dans le temps.

On étudie le cas d'un modèle de variation linéaire.

On estime simultanément les paramètres du modèle linéaire et le résultat de l'ACP par des processus d'approximation stochastique.

5. ACP D'UN VECTEUR ALEATOIRE (1)

Soit un vecteur aléatoire Z dans R^p.
R^p est muni d'une métrique M.
L'ACP du vecteur aléatoire Z consiste à :

1) rechercher une combinaison linéaire des composantes centrées de Z,
$f^{1\prime}(Z-E[Z])$, f^1 appartenant au dual R^{p*} de R^p,

. de variance maximale
. sous la contrainte de normalisation $f^{1\prime}M^{-1}f^1 = 1$;

2) rechercher une deuxième combinaison linéaire des composantes centrées de Z,
$f^{2\prime}(Z-E[Z])$,

. non corrélée à la première,
. de variance maximale
. sous la contrainte $f^{2\prime}M^{-1}f^2 = 1$;

3) et ainsi de suite jusqu'à un rang r au plus égal à p.

5. ACP D'UN VECTEUR ALEATOIRE (2)

La $i^{ème}$ combinaison linéaire est appelée le $i^{ème}$ facteur ;

on appelle également $i^{ème}$ facteur le vecteur $f^{\,i}$ des coefficients de la combinaison.

Soit

$$C = E[\,(Z - E(Z))(Z - E(Z))'\,] = E[\,ZZ'\,] - E[Z]\,E[Z'\,]$$

la matrice de covariance de Z.

$f^{\,i}$ est vecteur propre M^{-1}*-unitaire* de MC associé à la $i^{ème}$ plus grande valeur propre.

Si Z a un ensemble fini de N réalisations,

l'ACP de Z équivaut à l'ACP usuelle du tableau (N,p) des réalisations,

le poids de chaque réalisation étant défini par sa probabilité.

6. METHODE DE GRADIENT (1)

On suppose dans ce paragraphe que la matrice de covariance C et la métrique M sont connues.

La fonction

$$F(x) = \frac{< MCx, x >_{M^{-1}}}{< x, x >_{M^{-1}}}$$

est maximale pour $x = f^1$ et minimale pour $x = f^p$ de gradient

$$G(x) = \frac{2M^{-1}}{x' M^{-1} x} \left(MC - F(x) I \right) x.$$

Pour déterminer le premier facteur f^1, on peut utiliser un processus de gradient (X_n) défini récursivement par

$$X_{n+1} = X_n + a_n (MC - F(X_n) I) X_n .$$

6. METHODE DE GRADIENT (2)

Pour déterminer les *r* premiers facteurs, on peut utiliser le processus (X_n) suivant :

$$Y_{n+1}^{i} = X_n^i + a_n\left(MC - F(X_n^i)I\right)X_n^i \qquad (i = 1,2,\ldots,r),$$
$$X_{n+1} = orth_{M^{-1}}(Y_{n+1}).$$

On obtient $X_{n+1} = (X_{n+1}^1, \ldots, X_{n+1}^r)$ à partir de $Y_{n+1} = (Y_{n+1}^1, \ldots, Y_{n+1}^r)$

en faisant une orthogonalisation de Gram-Schmidt au sens de M^{-1}.

En supposant les *r* plus grandes valeurs propres de *MC* distinctes,

alors, pour $i = 1, \ldots, r$, le processus (X_n^i) converge en direction vers f^i,

en prenant la suite réelle (a_n) telle que

$$a_n > 0, \quad \sum_1^\infty a_n = \infty, \quad \sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty, \quad \sum_1^\infty a_n^2 < \infty.$$

7. APPROXIMATION STOCHASTIQUE DES FACTEURS D'UNE ACP (1)

Modèle

On a une suite d'observations de dimension p, $(Z_1,\ldots,Z_n,\ldots)$:

. $Z_n = \theta_n + R_n$,

. $(R_1,\ldots,R_n,\ldots)$ échantillon i.i.d. d'un vecteur aléatoire R, $E[R] = 0$, $Covar\,[R] = C$.

ACP partielle de Z = ACP de R

M métrique dans R^p

Les facteurs de l'ACP de R sont vecteurs propres de

$$B = MC = M\,(\,E[Z_n Z_n'\,] - \theta_n \theta_n'\,).$$

Lorsque $\theta_n = \theta$: ACP de R = ACP de Z.

Estimateurs de M et θ_n

Soit, au temps n,

M_n un estimateur de M et Θ_n un estimateur de θ_n ,fonctions de $Z_1,\ldots, Z_{n-1}$ et calculés récursivement.

7. APPROXIMATION STOCHASTIQUE DES FACTEURS D'UNE ACP (2)

Soit

$$B_n = M_n(Z_n Z_n' - \Theta_n \Theta_n'), \quad F_n(X_n^i) = \frac{< B_n X_n^i, X_n^i >_{M_n^{-1}}}{< X_n^i, X_n^i >_{M_n^{-1}}}.$$

Processus d'approximation stochastique des *r* premiers facteurs

$$pour\ i = 1,2,\ldots,r,$$

$$Y_{n+1}^i = X_n^i + a_n \left(B_n - F_n(X_n^i) I\right) X_n^i.$$

$$(X_{n+1}^1, \ldots, X_{n+1}^r) = orth_{M_n^{-1}}\left(Y_{n+1}^1, \ldots, Y_{n+1}^r\right)$$

Remarque. Si le calcul de M_n^{-1} nécessite trop de temps ou de mémoire, on peut être amené à définir un estimateur récursif de M^{-1}.

7. APPROXIMATION STOCHASTIQUE DES FACTEURS D'UNE ACP (3)

Convergence

(X_n^i) converge p.s. vers $f^{\,i}$, $i = 1,\ldots,r$, sous les hypothèses :

(1) R admet des moments d'ordre $4r$

(2) $\sup_n \|\theta_n\| < \infty$

(3) $M_n \quad T_n - mesurable\,;\, M_n \xrightarrow{p.s.} M \;;\, \sum_1^{\infty} a_n \left\| M_n - M \right\| < \infty \;\; p.s.$

(4) $\Theta_n \quad T_n - mesurable\,;\, \Theta_n - \theta_n \xrightarrow{p.s.} 0\,;\, \sum_1^{\infty} a_n \left\| \Theta_n - \theta_n \right\| < \infty \;\; p.s.$

(5) $a_n > 0, \sum_1^{\infty} a_n = \infty, \sum_1^{\infty} a_n^2 < \infty$

8. VARIANTES

1) Au pas *n*, on peut utiliser plusieurs observations $Z_{n1},\dots,Z_{nm_n}$ de *Z*.

On définit alors :

$$B_n = M_n\left(\frac{1}{m_n}\sum_{k=1}^{m_n} Z_{nk}Z_{nk}' - \Theta_n\Theta_n'\right).$$

2) Au pas *n*, on peut utiliser toutes les observations de *Z* faites jusqu'à ce pas.

On définit alors :

$$B_n = M_n\left(\frac{1}{n}\sum_{i=1}^{n} Z_iZ_i' - \Theta_n\Theta_n'\right);$$

B_n est un estimateur d'ordre *n* de $B = MC$.

9. ANALYSE FACTORIELLE D'UN VECTEUR ALEATOIRE «MULTIPLE »

Modèle

$(Z_1,\ldots,Z_n)$ suite de vecteurs aléatoires dans R^p observés :

. $Z_n = \theta + R_n$,
. $(R_1,\ldots,R_n)$ échantillon i.i.d. de R, $E[R] = 0$, $Covar\,[R] = C$.

L'ensemble des composantes de Z est partitionné en q groupes de variables aléatoires réelles

$$\{Z^{k1},\ldots,Z^{km_k}\}, k = 1,\ldots,q\ ;\ m_1 + \ldots + m_q = p.$$

$$Z^k = \left(Z^{k1},\ldots,Z^{km_k}\right)\quad v.a.\quad dans\quad R^{m_k},\ k = 1,\ldots,q.$$

$$Z = \left(Z^1,\ldots,Z^q\right): vecteur\ "multiple".$$

But

Réaliser une ACP de Z dans laquelle les vecteurs aléatoires Z^k aient un rôle équilibré :

on veut éviter que les premiers facteurs soient principalement déterminés à partir de certains Z^k.

10. SOLUTION DE L'ANALYSE CANONIQUE GENERALISEE (1)

Métrique *M* dans R^p

ACG de Z = ACP de Z dans R^p muni de la métrique

$$M = \begin{pmatrix} (C^1)^{-1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & (C^q)^{-1} \end{pmatrix}$$

C^k : matrice de covariance de Z^k.

Estimateur de C^k

On peut utiliser au pas n la matrice de covariance empirique N_n^k calculée récursivement à partir de $Z_1,\ldots,Z_{n-1}$.

Estimateur de M^{-1}

$$N_n = \begin{pmatrix} N_n^1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & N_n^q \end{pmatrix}$$

10. SOLUTION DE L'ANALYSE CANONIQUE GENERALISEE (2)

Estimateur de $(C^k)^{-1}$

a) On peut utiliser au pas n l'inverse M_n^k de N_n^k que l'on peut calculer récursivement.

b) Soit $Z_1^k = ((Z^k)' \quad 1)'$, vecteur aléatoire $(m_k+1, 1)$,
$J = (I \quad 0)'$ matrice (m_k+1, m_k).

$$X^k = \begin{pmatrix} (C^k)^{-1} \\ -E[Z^k{}'](C^k)^{-1} \end{pmatrix}$$ est la solution du système en X :

$$E[Z_1^k Z_1^k{}'] X - J = 0.$$

Processus d'approximation stochastique de X^k dans l'ensemble des matrices (m_k+1, m_k) :

$$M_{1,n+1}^k = M_{1n}^k - a_n \left(Z_{1n}^k Z_{1n}^k{}' M_{1n}^k - J \right).$$

10. SOLUTION DE L'ANALYSE CANONIQUE GENERALISEE (3)

En enlevant pour tout n la dernière ligne de $M_{1n}{}^{k}$, on obtient un processus $(M_n{}^k)$ qui converge presque sûrement vers $(C^k)^{-1}$.

Estimateur de M

$$M_n = \begin{pmatrix} M_n^1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & M_n^q \end{pmatrix}.$$

11. SOLUTION DE L'ANALYSE FACTORIELLE MULTIPLE (1)

Métrique M dans R^p

1) Pour $k = 1, 2, \ldots, q$:

On effectue l'ACP de Z^k dans R^{m_k} muni d'une métrique M^k ;

λ_1^k : plus grande valeur propre de $B^k = M^k C^k$.

2) On effectue l'ACP de Z dans R^p muni de la métrique

$$M = \begin{pmatrix} \frac{M^1}{\lambda_1^1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \frac{M^q}{\lambda_1^q} \end{pmatrix}.$$

3) Processus (M_n^k) d'estimation de M^k

Par exemple :

M^k métrique de l'inverse des variances des composantes de Z^k ;

M_n^k métrique de l'inverse des variances empiriques calculées récursivement à partir de $Z_1^k, \ldots, Z_{n-1}^k$.

11. SOLUTION DE L'ANALYSE FACTORIELLE MULTIPLE (2)

Processus (V_n^k) d'approximation stochastique d'un vecteur propre associé à la plus grande valeur propre λ_1^k de $B^k = M^k C^k$

$$\Theta_n^k = \frac{1}{n-1}\sum_{i=1}^{n-1} Z_i^k$$

$$B_n^k = M_n^k\left(Z_n^k Z_n^k{}' - \Theta_n^k \Theta_n^k{}'\right)$$

$$F_n^k(V_n^k) = \frac{\left\langle B_n^k V_n^k, V_n^k \right\rangle_{\left(M_n^k\right)^{-1}}}{\left\langle V_n^k, V_n^k \right\rangle_{\left(M_n^k\right)^{-1}}}$$

$$V_{n+1}^k = V_n^k + a_n\left(B_n^k - F_n^k\left(V_n^k\right)I\right)V_n^k.$$

Processus (Λ_n^k) d'estimation de λ_1^k

$$W_n^k = M_n^k\left(\frac{1}{n-1}\sum_{i=1}^{n-1} Z_i^k Z_i^k{}' - \Theta_n^k \Theta_n^k{}'\right), \textit{ estimateur de } B^k$$

$$\Lambda_n^k = \frac{\left\langle W_n^k V_n^k, V_n^k \right\rangle_{\left(M_n^k\right)^{-1}}}{\left\langle V_n^k, V_n^k \right\rangle_{\left(M_n^k\right)^{-1}}}$$

11. SOLUTION DE L'ANALYSE FACTORIELLE MULTIPLE (3)

Processus *(M_n)* d'estimation de *M*

$$M_n = \begin{pmatrix} \frac{M_n^1}{\Lambda_n^1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \frac{M_n^q}{\Lambda_n^q} \end{pmatrix}$$

Hypothèses de convergence des processus *(X_n^i)*

Hypothèses (1) et (5) du cas général et

$$(3')\ Pour\ k=1,\dots,q : M_n^k\ T_n - mesurable,\ M_n^k \xrightarrow{p.s.} M^k, \sum_1^\infty a_n \left\| M_n^k - M^k \right\| < \infty\ p.s.$$

$$(5')\ \sum_1^\infty \frac{a_n}{\sqrt{n}} < \infty.$$

12. ACP PARTIELLE D'UN VECTEUR ALEATOIRE (1)

Modèle

On a une suite d'observations $(Z_1,\ldots,Z_n,\ldots)$:

. $E[Z_n] = \theta_n$ dépend de n,

. les vecteurs $R_n = Z_n - E[Z_n]$ constituent un échantillon i.i.d. d'un vecteur aléatoire R de matrice de covariance C.

ACP partielle de Z = ACP de R

Les facteurs de l'ACP de R sont vecteurs propres de MC.

Modèle linéaire de variation de l'espérance de Z_n

Soit $\theta_n^1,\ldots,\theta_n^p$ les composantes de $\theta_n = E[Z_n]$. On suppose que pour tout k,

$$\theta_n^k = \left(U_n^k\right)'\beta^k,\ \beta^k \in R^{q_k},\ U_n^k \in R^{q_k}\ ;$$

β^k est un vecteur inconnu et U_n^k un vecteur connu au temps n à q_k composantes.

12. ACP PARTIELLE D'UN VECTEUR ALEATOIRE (2)

$$Z_n^k = \left(U_n^k\right)' \beta^k + R_n^k, \; k = 1, \ldots, p \; .$$

U_n^k peut être :

. un vecteur de fonctions connues du temps,

. un vecteur de valeurs de variables explicatives.

Estimation de β^k

On utilise le processus d'approximation stochastique de type Robbins-Monro (B_n^k) défini récursivement par

$$B_{n+1}^k = B_n^k - a_n U_n^k (U_n^k{}' B_n^k - Z_n^k).$$

12. ACP PARTIELLE D'UN VECTEUR ALEATOIRE (3)

Pour tout $k = 1,\ldots,p$, ce processus converge presque sûrement vers β^k sous les hypothèses

$$.\sup_n \left\| U_n^k \right\| < \infty,$$

$$.\exists\, r_k \in N,\ \exists\, \lambda_k > 0,\ \exists \left(n_{kl}, l \geq 1\right), n_{k1} = 1,\ n_{kl} < n_{k,l+1} \leq n_{kl} + r_k,$$

$$\lambda_{\min}\left(\sum_{j \in I_{kl}} U_j^k U_j^{k\,\prime} \right) \geq \lambda_k,\ I_{kl} = \left\{ n_{kl}, \ldots, n_{k,l+1} - 1 \right\},$$

$$.a_n > 0,\ \sum_1^\infty \min_{j \in I_{kl}} a_j = \infty,\ \sum_1^\infty a_n^2 < \infty.$$

12. ACP PARTIELLE D'UN VECTEUR ALEATOIRE (4)

On définit un estimateur de θ_n^k : $\Theta_n^k = U_n^k{}' B_n^k,$

et un estimateur de $\theta_n = E[Z_n]$: $\Theta_n = \left(\Theta_n^1, \ldots, \Theta_n^p\right)'.$

Soit, au temps n, M_n un estimateur de M fonction de $Z_1, \ldots, Z_{n-1}$. Soit

$$B_n = M_n(Z_n Z_n' - \Theta_n \Theta_n'), \quad F_n(X_n^i) = \frac{< B_n X_n^i, X_n^i >_{M_n^{-1}}}{< X_n^i, X_n^i >_{M_n^{-1}}}.$$

On définit le processus d'approximation stochastique (X_n) tel que :

$$pour\ i = 1,2,\ldots,r,$$

$$Y_{n+1}^i = X_n^i + a_n\left(B_n - F_n(X_n^i) I\right) X_n^i.$$

$$\left(X_{n+1}^1, \ldots, X_{n+1}^r\right) = orth_{M_n^{-1}}\left(Y_{n+1}^1, \ldots, Y_{n+1}^r\right)$$

12. ACP PARTIELLE D'UN VECTEUR ALEATOIRE (5)

Le processus *(X_n)* converge presque sûrement en direction vers le *i*ème facteur de l'ACP de *R* sous :

. les hypothèses énoncées relatives à *(U_n^k)*,

. les hypothèses énoncées relatives à *(M_n)*,

. et en prenant $a_n = \frac{c}{n^{\alpha}},\ \ c>0,\ \ \frac{1}{2}<\alpha\leq 1.$

CAS DE L'ACP NORMEE PARTIELLE

M est la métrique diagonale de l'inverse des variances des composantes de *R*.

On définit le *k*ième élément diagonal de M_n, M_n^k, estimateur au temps *n* de l'inverse de la variance de R^k, que l'on peut déterminer récursivement, par :

$$\frac{1}{M_n^k} = \frac{1}{n-1}\sum_{j=1}^{n-1}\left(Z_j^k - \Theta_j^k\right)^2 .$$

13. ACG PARTIELLE D'UN VECTEUR ALEATOIRE (1)

Modèle

$(Z_1,\ldots,Z_n,\ldots)$ suite de vecteurs aléatoires réels de dimension p :

$Z_n = (Z_n^1,\ldots,Z_n^p)$; $Z_n^k = \theta_n^k + R_n^k = (U_n^k)'\beta^k + R_n^k$, $k = 1,\ldots,p$

$R_n = (R_n^1,\ldots,R_n^p)$

$(R_1,\ldots,R_n,\ldots)$ échantillon i.i.d. de R, v.a. de dim p, $E[R] = 0$, $Covar\,[R] = C$

$R = (R^1,\ldots,R^q)$, R^k vecteur aléatoire de dimension m_k,

$E[R^k] = 0$, $Covar\,[R^k] = C^k$.

ACG partielle de Z

ACP de R avec la métrique

$$M = \begin{pmatrix} \left(C^1\right)^{-1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \left(C^q\right)^{-1} \end{pmatrix}.$$

Les facteurs $f^{\,i}$ sont vecteurs propres de MC.

Les vecteurs $u^{\,i} = M^{-1}f^{\,i}$, vecteurs unitaires des axes principaux, deux à deux M-orthogonaux, sont vecteurs propres de $CM = (\,E[ZZ'] - E[Z]E[Z']\,)\,M$.

13. ACG PARTIELLE D'UN VECTEUR ALEATOIRE (2)

Processus (A_n^k) d'approximation stochastique de $(C^k)^{-1}$

$C^k = Covar\,[R^k] = Covar\,[Z_n^k] = E\,[\,Z_n^k\,(Z_n^k - E[Z_n^k])'\,]$

$(C^k)^{-1}$ est solution pour tout n de l'équation en A, $E[Z_n^k\,(Z_n^k - E[Z_n^k])'\,]\,A = I$.

$$A_{n+1}^k = A_n^k - a_n\left(Z_n^k\left(Z_n^k - \Theta_n^k\right)' A_n^k - I\right)$$

Processus (M_n) d'approximation stochastique de M

$$M_n = \begin{pmatrix} A_n^1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & A_n^q \end{pmatrix}$$

Processus (X_n^i) d'approximation stochastique de u^i

Processus général avec $B_n = (Z_n Z_n' - \Theta_n \Theta_n')\,M_n$.

14. EXTENSIONS

Modèle linéaire généralisé

$$Z_n^k = h^k(\,(U_n^k)'\beta^k\,) + R_n^k\;,\; k = 1,\dots,p.$$

Variation de la variance de Z_n^k

$$Z_n^k = (U_n^k)'\,\beta^k + \delta_n^k\,R_n^k\;,$$
$$(\delta_n^k)^2 = (V_n^k)'\delta^k,\; k = 1,\dots,p.$$

15. ACP PROJETEE (1)

Soit un couple de vecteurs aléatoires réels (R,S), dim $R = p$, dim $S = q$.

ACP projetée de S par rapport à R = ACP de $E[S|R]$.

Modèle : $E[S|R] = A'R + D$, A (p,q), D (q,1).

1) CAS OU IL N'EXISTE PAS DE RELATION AFFINE ENTRE LES COMPOSANTES de R

Alors : $A = (Covar\,[R])^{-1} Covar\,[R,S]$

Soit $R_1 = (R' \;\; 1)'$, $A_1 = (A' \;\; D)'$.

A_1 est la solution unique du système en X :

$$E[R_1 R_1'\,]\, X = E[R_1 S'\,].$$

Soit $((R_n,S_n))$ un échantillon i.i.d. de (R,S) et $R_{1n} = (R_n' \;\; 1)'$.

Un processus d'approximation stochastique de A_1 est (A_{1n}) :

$$A_{1,n+1} = A_{1n} - a_n\,(R_{1n} R_{1n}' A_{1n} - R_{1n} S_n'\,).$$

En enlevant pour tout n la dernière ligne de A_{1n}, on obtient le processus (A_n) d'approximation stochastique de A.

15. ACP PROJETEEE (2)

Approximation stochastique des facteurs de l'ACP projetée

On définit une métrique M dans R^q.

Les facteurs de l'ACP de $E[S|R]$ sont vecteurs propres de

$B = M \text{ Covar } [\, E[R|S]\,] = M \text{ Covar } [S,R]\, A = M\, (\, E[SR'\,] - E[S]\, E[R'\,]\,)\, A.$

Soit (M_n) un estimateur de M.

Soit $B_n = M_n \left(S_n R_n' - \overline{S_{n-1}}\ \overline{R_{n-1}}' \right) A_n$.

On définit le processus $(X_n) = (X_n^1, \ldots, X_n^r)$:

$$Y_{n+1}^i = X_n^i + a_n \left(B_n - F_n(X_n^i) I \right) X_n^i \ , \ i = 1, \ldots, r$$
$$\left(X_{n+1}^1, \ldots, X_{n+1}^r \right) = orth_{M_n^{-1}} \left(Y_{n+1}^1, \ldots, Y_{n+1}^r \right)$$

Cas particulier

$M = (\text{Covar } [S])^{-1}$: *analyse canonique des vecteurs aléatoires R et S*.

$B = (\text{Covar } [S])^{-1} \text{ Covar } [S,R]\, A = FA$.

On définit un processus (F_n) d'approximation stochastique de F, semblable à (A_n).

Soit $B_n = F_n A_n$.

On définit le processus (X_n) semblable au précédent.

15. ACP PROJETEE (3)

2) CAS OU LES COMPOSANTES DE *R* SONT LES INDICATRICES DES MODALITES EXCLUSIVES D'UNE VARIABLE ALEATOIRE NOMINALE

Dans ce cas : $E[S|R] = A'R$

avec : $A = (E[RR'])^{-1} Covar\,[R,S] = A_2 - u\,E[S']$,

$A_2 = (E[RR'])^{-1} E[RS']$, $u = (11\ldots1)'$.

Les facteurs de l'ACP projetée sont vecteurs propres de $M\ Covar\,[S,R]\ A_2$.

On définit un processus (A_{2n}) d'approximation stochastique de A_2 :

$$A_{2,n+1} = A_{2n} - a_n\,(R_n R_n' A_{2n} - R_n S_n').$$

Soit $B_n = M_n \left(S_n R_n' - \overline{S_{n-1}}\ \overline{R_{n-1}}' \right) A_{2n}.$

On définit un processus (X_n) d'approximation stochastique des facteurs semblable au précédent.

Cas particuliers

. $M = (Covar\,[S])^{-1}$: <u>*analyse factorielle discriminante*</u>.

. Les composantes de *S* sont les indicatrices des modalités exclusives d'une variable aléatoire nominale ; $M = (E[SS'])^{-1}$: <u>*analyse factorielle des correspondances*</u>.

On peut définir dans ces deux cas un processus spécifique.

REFERENCES CITEES

[ALB 67] Albert A.E., Gardner L.A., « Stochastic approximation and non linear regression », Research Monograph 42, The M.I.T. Press, 1967.

[BEN 69] Benzecri J.P., « Approximation stochastique dans une algèbre normée non commutative », *Bulletin de la SMF*, vol. 97, p. 225-241, 1969.

[BOU 96] Bouamaine A., Thèse de Doctorat d'Etat ès Sciences Appliquées, EMI, Rabat, 1996.

[BOU 97] Bouamaine A., Monnez J.M., « Convergence d'une classe de processus d'approximation stochastique », *Publications de l'ISUP*, vol. 41, n°1-2, p. 97-117, 1997.

[BOU 98] Bouamaine A., Monnez J.M., « Approximation stochastique de vecteurs et valeurs propres », *Publications de l'ISUP*, vol. 42, n°2-3, p. 15-38, 1998.

[KRA 70] Krasulina T.P., "Method of stochastic approximation in the determination of the largest eigenvalue of the mathematical expectation of random matrices", *Automation and Remote Control*, vol. 2, p. 215-221, 1970.

[MAC 67] MacQueen J., «Some methods for classification and analysis of multivariate observations », Proceedings of the 5th Berkeley Symposium on Probability and Statistics, p. 281-297, 1967.

[MON 94] Monnez J.M., « Convergence d'un processus d'approximation stochastique en analyse factorielle », *Publications de l'ISUP*, vol. 38, n°1, p.37-56, 1994.

[MON 06a] Monnez J.M., « Almost sure convergence of stochastic gradient processes with matrix step sizes », *Statistics & Probability Letters,* vol. 76, n°5, p.531-536, 2006.

[MON 06b] Monnez J.M., « Approximation stochastique en analyse factorielle multiple », *Publications de l'ISUP*, vol. 50, n°3, p. 27-45, 2006.

[MON 08a] Monnez J.M., "Stochastic approximation of the factors of a generalized canonical correlation analysis", *Statistics & Probability Letters,* vol. 78, n°14, p. 2210-2216, 2008.

[MON 08b] Monnez J.M., « Analyse en composantes principales d'un flux de données d'espérance variable dans le temps », *Revue des Nouvelles Technologies de l'Information,* Vol C-2, p. 43-56, 2008.