

M1 MMSES

Outils pour la modélisation stochastique

Gabriel Lang
UMR 518 MIA, AgroParisTech-INRA

10 octobre 2011

1 Introduction

Ce document présente un ensemble de modèles aléatoires couramment utilisés. Il est construit pour servir de repère avant de se reporter à des ouvrages spécialisés sur des modèles particuliers. La notion principale introduite est la notion de dépendance entre variables aléatoires, qui confère une structure aux phénomènes aléatoires. Les modèles présentés ici sont intermédiaires entre deux archétypes : le modèle déterministe où toutes les conditions sont contrôlées et les prévisions certaines et le modèle de hasard parfait où aucune prévision n'est possible comme dans les tirages successifs d'un dé.

2 Prérequis

La lecture de ce document nécessite d'être familiarisé avec les notions élémentaires d'algèbre linéaire, de probabilité et de statistiques. En algèbre linéaire, doivent être connus les applications linéaires, les opérations matricielles ainsi que les formes quadratiques positives et les résultats de diagonalisation des matrices symétriques. En probabilité, il est nécessaire de connaître les notions d'espace probabilisé, de variable aléatoire, de loi de probabilité discrète et continue d'une variable et d'un couple de variables, de calculs de moments et d'indépendance de variables aléatoires. Pour la partie statistique, une connaissance élémentaire de l'estimation et de la régression linéaire est suffisante.

3 Rappels

3.1 Variable aléatoire

On rappelle qu'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ se compose d'un ensemble Ω , d'un ensemble \mathcal{A} (appelé tribu) de parties de Ω appelées événements, stable par intersection, union infinie et complémentarité et d'une mesure positive \mathbb{P} de masse 1.

Définition 1. Une variable aléatoire réelle X est une fonction mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B})$ où \mathcal{B} est la tribu de \mathbb{R} engendrée par les intervalles ouverts. Une fonction f est dite mesurable si toute image réciproque par f d'une partie de \mathbb{R} appartenant à \mathcal{B} appartient à \mathcal{A} .

Une variable aléatoire X est définie comme une fonction. Mais elle est surtout utilisée pour définir une loi de probabilité sur \mathbb{R} par transport de Ω sur \mathbb{R} . On définit une probabilité \mathbb{P}' sur les intervalles I de \mathbb{R} par $\mathbb{P}'(I) = \mathbb{P}(X^{-1}(I)) = \mathbb{P}\{\omega; X(\omega) \in I\}$.

3.2 Dépendance des variables aléatoires

Deux variables aléatoires réelles X et Y sont indépendantes si et seulement si pour tout intervalle A et B de \mathbb{R} :

$$P(X \in A \text{ et } Y \in B) = P(X \in A)P(Y \in B).$$

La dépendance de deux variables se mesure par la covariance:

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Propriétés:

- Pour des variables d'espérance nulle, la covariance est un produit scalaire. Deux variables seront dites orthogonales si leur covariance est nulle.
- Deux variables indépendantes sont orthogonales.
- Deux variables orthogonales ne sont pas toujours indépendantes (même si ce sont des variables de loi gaussienne).

Soit $X = (X_1, \dots, X_p)$ un vecteur aléatoire ; on associe à X la matrice de covariance Σ formée des $\text{cov}(X_i, X_j)$ pour i et j variant de 1 à p . Cette matrice est symétrique et positive. Si (a_1, \dots, a_p) et (b_1, \dots, b_p) sont des vecteurs réels :

$$\text{cov}(a_1X_1 + \dots + a_pX_p, b_1X_1 + \dots + b_pX_p) = (a_1, \dots, a_p)\Sigma^t(b_1, \dots, b_p).$$

Si P est la matrice d'une transformation linéaire et $Y = PX$, la matrice de covariance de Y est $P\Sigma^tP$.

4 Lois de probabilité

Pour la raison précédente, la définition d'une variable aléatoire est souvent réduite à la donnée de la loi de probabilité sur \mathbb{R} qu'elle induit. Parmi les lois de probabilité sur \mathbb{R} , on distingue les lois discrètes dont le support se réduit à un ensemble dénombrable de points de \mathbb{R} et les lois de probabilité continues qui donnent une probabilité nulle à tout ensemble ne contenant pas un intervalle ouvert de \mathbb{R} . Les lois discrètes sont définies en donnant la probabilité de tous les points du support. Les lois de probabilités continues (sous-entendu par rapport à la mesure de Lebesgue sur \mathbb{R}) se définissent par leur densité par rapport à cette mesure. Il existe bien sûr des cas mixtes ainsi que des lois de probabilités dont les supports ne sont ni des intervalles ni des points de \mathbb{R} . Nous rappelons les deux lois de probabilité très courantes que nous utiliserons le plus par la suite.

4.1 Loi gaussienne $\mathcal{N}(\mu, \sigma^2)$

La loi gaussienne ou normale est la loi de probabilité continue sur \mathbb{R} définie par la densité :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

où $\mu \in \mathbb{R}$ et $\sigma > 0$. Tous les moments de la loi sont finis

$$\int_{-\infty}^{+\infty} x^p f(x) dx < +\infty,$$

pour tout $p \in \mathbb{N}$. En particulier, l'espérance vaut μ et la variance σ^2 .

4.2 Loi de Poisson $\mathcal{P}(\theta)$

La loi de Poisson est une loi discrète définie sur \mathbb{N} par :

$$\mathbb{P}(X = k) = \exp(-\theta) \frac{\theta^k}{k!}$$

où $\theta > 0$. Tous les moments de la loi sont finis:

$$\sum_{k=0}^{\infty} k^p P(X = k) < \infty.$$

En particulier l'espérance et la variance valent θ .

4.3 Stabilité de ces lois

Ces lois sont stables au sens suivant : si X_1 et X_2 sont deux variables indépendantes de loi gaussienne, leur somme est une variable de loi gaussienne. De plus, si a est un réel aX_1 est une variable de loi gaussienne. Si X_1 et X_2 sont deux variables indépendantes de loi de Poisson, leur somme est une variable de loi de Poisson. Par contre, si a est un réel aX_1 n'est pas une variable de loi de Poisson, car le support de la loi n'est plus nécessairement l'ensemble des nombres entiers. La stabilité de la loi normale entraîne le résultat suivant :

Théorème 1. Théorème de la limite centrale :

Soit une collection de n variables (X_1, \dots, X_n) indépendantes de même loi, d'espérance nulle et de variance finie σ^2 . Soit S_n la somme de ces variables. Alors la loi de S_n/\sqrt{n} tend vers la loi $\mathcal{N}(0, \sigma^2)$.

Ce résultat s'interprète de la façon suivante : si un phénomène est le résultat de l'addition d'une multitude de petites variations aléatoires indépendantes de même variabilité, ce phénomène suit une loi normale. Il peut s'agir de la taille d'un individu dans une population, d'un cours boursier résultat d'un grand nombre d'ordre d'achat et de ventes...

5 Vecteur aléatoire gaussien

Soit p un entier et (X_1, \dots, X_p) , p variables aléatoires.

Définition 2. (X_1, \dots, X_p) forment un vecteur aléatoire gaussien si et seulement si pour tout vecteur réel (a_1, \dots, a_p) , la variable $a_1X_1 + \dots + a_pX_p$ suit une loi gaussienne.

La notion de vecteur gaussien définit non seulement la loi des variables coordonnées mais également la relation de dépendance entre ces variables :

- Les variables coordonnées ont toutes une loi gaussienne.
- Un vecteur formé de variables coordonnées gaussiennes n'est pas nécessairement gaussien.
- Un vecteur formé de variables coordonnées gaussiennes orthogonales n'est pas nécessairement gaussien.

- Un vecteur formé de variables coordonnées gaussiennes indépendantes est gaussien.
- Si dans un vecteur gaussien, les coordonnées sont orthogonales, alors elles sont indépendantes.
- Le vecteur image d'un vecteur gaussien par une transformation linéaire est gaussien.

Tout vecteur gaussien X peut être écrit comme la transformation linéaire d'un vecteur de variables gaussiennes indépendantes. Supposons pour simplifier que les espérances des variables coordonnées sont nulles ; la matrice Σ est réelle et symétrique, donc on peut trouver une matrice P orthonormale (${}^tP = P^{-1}$) et une matrice diagonale D telles que $\Sigma = {}^tPDP$. Posons $Y = {}^tPX$. Le vecteur Y est un vecteur gaussien et sa matrice de covariance est D qui est diagonale. Les coordonnées de Y sont orthogonales et donc indépendantes.

6 Les chaînes de Markov

Lorsque l'on observe une série chronologique de données réelles, la question de la prédiction vient naturellement : dans quelle mesure la donnée actuelle me renseigne-t-elle sur la prochaine valeur observée ? Nous avons vu précédemment qu'un processus peut être défini par la donnée de ses lois marginales finies. Mais il existe d'autre façon de définir des processus. Nous avons vu par exemple le cas du processus autorégressif où chaque variable était définie à partir de la précédente et d'un bruit aléatoire appelé innovation. On peut définir un processus par la donnée de lois conditionnelles d'une variable en fonction des précédentes. On définira :

$\mathcal{L}(X_0)$, la loi de X_0 ; $\mathcal{L}(X_1|X_0)$, la loi de X_1 connaissant X_0 , $\mathcal{L}(X_2|X_1, X_0)$ la loi de X_2 connaissant X_1 et X_0 ... En utilisant la formule de Bayes, on déduit successivement de ces données les lois jointes de chaque n -uplet de variables du type (X_0, X_1, \dots, X_n) . La donnée de ces lois conditionnelles est donc équivalente à la donnée des lois marginales. Le nombre de paramètres est a priori aussi important. Pour définir des modèles utilisables, nous allons comme précédemment limiter le nombre de paramètres : la façon la plus simple est de considérer que les lois conditionnelles définissant une variable en fonction des précédentes ne sont pas fonction de l'ensemble des variables précédentes mais seulement de la dernière. Cette idée est la base de la définition des processus markoviens.

6.1 Définition

Comme précédemment, le processus est défini pour un espace de temps discret \mathbb{N} . L'espace d'état E est fini. Pour simplifier les notations, on numérote les états et $E = 1, 2, \dots, k$. Le processus X est une application de $\mathbb{N} \times \Omega$ dans E . Le processus possède la propriété de Markov si la loi de la variable aléatoire X_n ne dépend des variables précédentes que par X_{n-1} :

$$\mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}).$$

6.2 Formalisme matriciel

E étant fini, cette loi conditionnelle est définie par l'ensemble des $\mathbb{P}((X_n = j | X_{n-1} = i))$ pour toutes les valeurs de i et de j (probabilité de passer de i à j). Ces nombres peuvent être rangés dans une matrice P_n de dimension $k \times k$ dont la somme des coefficients par ligne est égale à 1. Les matrices de ce type sont appelées matrices stochastiques. De même, la loi marginale d'une

valeur X_n peut être notée sous la forme du vecteur L_n de dimension k dont les coordonnées sont les $P(X_n = i)$. On peut calculer la loi de X_n en fonction de la loi de X_{n-1} par la formule de Bayes :

$$\mathbb{P}(X_n = i) = \sum_{j=1}^k \mathbb{P}((X_n = i | X_{n-1} = j) \mathbb{P}(X_{n-1} = j).$$

Dans le formalisme matriciel choisi, cette formule se résume à :

$$L_n = L_{n-1} P_n.$$

On en déduit que $L_n = L_0 P_1 P_2 \cdots P_n$.

6.3 Chaînes de Markov homogènes

On considère le cas où la loi conditionnelle de X_n sachant X_{n-1} ne varie pas au cours du temps. Dans ce cas, la matrice P_n est une matrice constante P et :

$$L_n = L_0 P^n.$$

Le processus est entièrement défini par la donnée de la matrice P et du vecteur L_0 .

Remarque: Il ne faut pas confondre homogène et stationnaire. L'homogénéité implique que la transformation de la loi de X_n à celle de X_{n+1} est constante, mais ne signifie pas que les X_i ont la même loi.

6.4 Propriété des chaînes de Markov

Le processus de Markov homogène a un comportement aléatoire : à chaque étape la réalisation observée est tirée selon la loi \mathcal{L}_n correspondante. A ce titre, le processus ne se stabilise jamais sur une valeur donnée (sauf cas extrêmement particuliers) ne suit pas des cycles réguliers, comportements observés dans les systèmes dynamiques déterministes. Nous allons cependant observer certains comportements moyens, qui correspondent aux notions de cycles et de stabilisation vers l'équilibre observés dans les systèmes déterministes.

6.4.1 Transience et absorption

Définition 3. Un état j tel que $\mathbb{P}_{i,j} = \mathbb{P}(X_n = j | X_{n-1} = i) = 0$ pour tout $i \neq j$ est dit transient. Un état i tel que $\mathbb{P}_{i,j} = \mathbb{P}(X_n = j | X_{n-1} = i) = 0$ pour tout $j \neq i$ est dit absorbant.

Si l'état j est transient, une fois que la chaîne quitte cet état, elle ne revient plus jamais dessus. Si l'état i est absorbant, une fois que la chaîne passe par cet état, elle ne le quitte plus. Ces propriétés peuvent être définies pour des ensembles d'états. Un ensemble A tel que $\mathbb{P}(X_n \in A | X_{n-1} \notin A) = 0$ est dit transient, un ensemble A tel que $\mathbb{P}(X_n \notin A | X_{n-1} \in A) = 0$ est dit absorbant.

6.4.2 Irréductibilité

Une chaîne est dite irréductible lorsqu'il est toujours possible au cours de l'évolution de la chaîne de passer d'un état à un autre, en un nombre d'étape fini. Cette propriété se traduit par : Pour tout i et j appartenant à E , il existe un temps n tel que $\mathbb{P}_{i,j}^n = \mathbb{P}(X_n = j | X_0 = i) \neq 0$. Une chaîne irréductible ne contient ni élément transient, ni élément absorbant. Une condition suffisante d'irréductibilité est que P ne contienne aucun zéro.

6.4.3 Périodicité

Une chaîne est périodique si on peut trouver une partition de l'ensemble E en sous-ensembles E_1, \dots, E_d tels que $\mathbb{P}(X_n \in E_{i+1} | X_{n-1} \in E_i) = 1$, avec la convention que $E_{d+1} = E_1$. Les trajectoires ne sont pas à proprement parler périodiques, puisque partant d'une valeur dans E_i , la valeur suivante peut être choisie au hasard dans E_{i+1} , mais les trajectoires passent périodiquement dans les sous-ensembles de la partition.

6.4.4 Loi stationnaire

Une loi stationnaire pour la chaîne est une loi \mathcal{L} telle que si X_n suit la loi \mathcal{L} , alors X_{n+1} suit la loi \mathcal{L} . La loi L vérifie donc : $LP = L$. On remarque qu'il s'agit d'un vecteur propre de la matrice P (en toute rigueur de sa transposée) pour la valeur propre 1, ce qui permet de calculer explicitement cette loi par des méthodes d'algèbre linéaire.

Propriété 1. Si une chaîne est irréductible, il existe une unique loi stationnaire.

Notes :

- Cette propriété n'est pas vraie lorsque la chaîne a un ensemble d'état E dénombrable et non fini. On peut alors simplement dire qu'il existe au plus une loi stationnaire.
- Si la loi \mathcal{L}_0 est la loi stationnaire \mathcal{L} , alors toutes les variables X_n ont la même loi de façon évidente. En fait on a une propriété plus forte, puisque le processus est stationnaire.
- Le processus peut être stationnaire et périodique. Supposons pour simplifier une chaîne déterministe périodique à trois états. Cette chaîne n'est qu'un cas particulier des processus de Markov homogènes définis précédemment correspondant à la matrice P :

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (1)$$

La mesure stationnaire associée est la mesure équiprobable $L=(1/3,1/3,1/3)$. Si $L_0 = L$, le processus est bien stationnaire; les lois de X_n sont égales à L , les lois jointes de (X_n, X_{n+1}) sont définies par :

$$\mathbb{P}(X_n = 1, X_{n+1} = 2) = \mathbb{P}(X_n = 2, X_{n+1} = 3) = \mathbb{P}(X_n = 3, X_{n+1} = 1) = 1/3.$$

et on peut vérifier que toutes les marginales finies sont bien celles d'un processus stationnaire. Cependant chaque trajectoire est cyclique. Ce n'est que pour un groupe de trajectoires dont les départs sont tirés selon la loi L que l'on peut observer que les variables X_n successives sont de même loi. La stationnarité est une propriété du processus qui n'est pas observable sur une trajectoire.

6.5 Stationnarisation

Nous avons vu au paragraphe précédent qu'une loi stationnaire était telle que partant de cette loi comme loi initiale, le processus de Markov était stationnaire. La loi stationnaire est un point fixe pour la transformation par P . Elle est l'équivalent d'un point d'équilibre d'un système dynamique déterministe. Comme dans le cas d'un système dynamique déterministe, nous allons examiner si ce point d'équilibre est stable ou instable.

Propriété 2. Si une chaîne est irréductible et apériodique, quelle que soit la loi initiale \mathcal{L}_0 , la loi de X_n tend vers la loi stationnaire \mathcal{L} :

$$L_0 P^n \xrightarrow{n \rightarrow +\infty} L.$$

Cette convergence se fait à vitesse exponentielle :

$$\|L_0 P^n - L\|_\infty \leq C \rho^n.$$

où C est une constante positive et $0 < \rho < 1$.

Une chaîne irréductible et apériodique se stabilise très vite vers la loi stationnaire. La loi stationnaire est l'analogie d'un point d'équilibre stable exponentiellement attractif.

6.6 Ergodicité

Nous venons de voir que les chaînes irréductibles se stabilisaient très vite; la loi de la variable X_n tend très vite vers la loi stationnaire. Ce comportement est un comportement global du processus; il n'indique rien sur le comportement d'une trajectoire particulière. Nous avons vu avec le cas des processus déterministes périodiques que le fait que le processus soit stationnaire n'apparaissait pas sur une trajectoire unique, mais que cette stationnarité était un comportement global portant sur l'ensemble des trajectoires. Dans le cas où on observe une unique trajectoire, est-il possible de se rendre compte que le processus est stationnaire? La distribution des $X_i(\omega)$ renseigne-t-elle sur la loi stationnaire \mathcal{L} ? Le fait que les positions successives d'une trajectoire soient distribuées selon la loi stationnaire s'appelle ergodicité ; si un processus est ergodique alors pour une fonction réelle f , la moyenne des $f(X_n(\omega))$ est égale à l'espérance d'une variable $f(X)$ où X est une variable de loi \mathcal{L} . Cette propriété est souvent résumée par la phrase sibylline suivante: La moyenne dans le temps est égale à la moyenne dans l'espace. Une définition rigoureuse de l'ergodicité est donnée par la définition suivante :

Définition 4. Une chaîne de Markov est dite fortement ergodique si

- elle admet une probabilité invariante \mathcal{L} .
- quelle que soit la loi initiale \mathcal{L}_0 , presque pour tout ω :

$$\frac{1}{n} \sum_{i=0}^n f(X_i(\omega)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}_{\mathcal{L}}(f) = \sum_{j=1}^k L_j f(j).$$

L'ergodicité d'une chaîne implique que :

- l'évolution de la chaîne dépend très peu des conditions initiales ; la chaîne se stabilise très vite.
- une seule trajectoire suffit à renseigner sur la distribution stationnaire que le processus atteint très vite. La fréquence de passage de la trajectoire par les différents états converge vers la probabilité \mathcal{L} .

Propriété 3. Une chaîne de Markov irréductible et apériodique est fortement ergodique.

6.7 Conclusion

L'application principale des chaînes de Markov utilise la propriété d'ergodicité forte. Dans certains problèmes de simulation, on veut tirer des échantillons d'une loi dont on ne connaît ni la densité, ni les densités marginales, mais les densités des lois conditionnelles. On construit une chaîne de Markov à partir de la connaissance de ces lois conditionnelles, dont on sait que la loi stationnaire est la loi qu'on cherche à simuler. On lance la chaîne de Markov à partir d'un point quelconque, et on commence à l'observer à partir d'un certain nombre d'itération : la chaîne s'étant stabilisée, les tirages observés forment un échantillon de la loi stationnaire.

7 Modèles pour des données réparties dans le temps

Nous nous intéressons à des ensembles de données qui ne sont pas la répétition d'une même mesure, mais des mesures prises à différents moments ou en différents endroits. Les valeurs de ces mesures seront par la suite notées $X(t)$, où t est l'indice représentant le lieu ou l'instant de la mesure. L'approche la plus simple consiste à dire qu'il existe un phénomène sous-jacent, déterministe, qui peut se décrire par une fonction simple $f(t)$ et que la mesure est le résultat de la composition (en général une simple addition) de cette fonction et d'un bruit de mesure aléatoire ε_t :

$$X(t) = f(t) + \varepsilon_t. \quad (2)$$

7.1 Prétraitement des données

Le premier traitement des données brutes va consister à proposer une fonction f déterministe pour la décomposition (2). La fonction f proposée classiquement se compose de deux parties. La première partie appelée tendance est une fonction d'une famille simple décrite par peu de paramètres : fonction affine (droite) , polynôme de degré faible, ou fonction exponentielle. La deuxième partie est une fonction périodique, appelée saisonnalité de la série. Elle apparaît naturellement lorsque le phénomène à modéliser possède une périodicité connue (annuelle pour une série météorologique ou certaines séries économiques). Le choix des familles de fonctions parmi lesquelles on cherche la fonction f se fait souvent à vue, en regardant les graphiques des données sans autre justification théorique : certaines saisonnalités sont visibles, d'autres apparaissent plus nettement après certaines transformations des données brutes (transformée de Fourier appelée périodogramme). Une tendance croissante régulière sera modélisée par une tendance affine, par un polynôme d'ordre 2 si on observe un certain creusement... Il n'existe pas de critère objectif pour faire le choix de cette famille. Ce choix fait, les méthodes pour trouver la meilleure fonction f dans la famille considérée sont souvent dérivées des méthodes de régression multiple.

7.2 Tendance polynomiale

On effectue une régression du vecteur $X(t)$ sur les fonctions puissance $(1, t, t^2, \dots)$ jusqu'au degré choisi, selon le modèle

$$X(t) = a_0 + a_1 t + \dots + a_p t^p + \varepsilon_t.$$

Le résultat de la régression

$$f(t) = \hat{a}_0 + \hat{a}_1 t + \dots + \hat{a}_p t^p$$

est la partie fonction déterministe et la différence $\eta_t = X(t) - f(t)$ est interprétée comme le bruit de mesure ; dans un modèle de régression classique, ce bruit est formé de variables indépendantes.

7.3 Tendance exponentielle

Si le modèle de tendance est logarithmique :

$$X(t) = A \exp(\lambda t) + \eta_t,$$

on transforme préalablement les données par la fonction logarithme

$$\log(X(t)) = \log(A) + \lambda t + \log(1 + \eta_t/A \exp(\lambda t)).$$

On effectue alors une régression linéaire (sans tenir compte du fait que le bruit indépendant n'est pas de variance constante).

7.4 Saisonalité

Si de plus on veut estimer une saisonnalité de période T choisie, on ajoute dans le modèle de régression polynômial un certain nombre de fonctions périodiques. Le choix des fonctions sur lesquelles on effectue la régression peut se faire de deux façons. On peut choisir la base des fonctions sinus:

$$h_k(t) = \sin(2\pi k/T), k = 1, \dots, T - 1$$

et ne conserver dans la saisonnalité que les coefficients de régression importants. On peut également choisir la base des fonctions indicatrices périodiques pour $k = 1, \dots, T - 1$:

$$Z_k(t) = 1 \text{ si } t - k \text{ est un multiple de } T, Z_k(t) = 0 \text{ sinon.}$$

Dans ce cas, les coefficients de régression ont une interprétation simple. Supposons que les données soit mensuelles et que la période T soit donc 12. Les coefficients de régression correspondent aux moyennes interannuelles pour les différents mois.

Dans les deux cas, pour que les vecteurs de régression ne soient pas liés linéairement, on n'utilise pas la fonction de régression correspondant à $k = T$. La série à laquelle on a retiré une saisonnalité estimée est baptisée série corrigée des variations saisonnières.

7.5 Modèle pour la partie aléatoire

Après soustraction de la partie déterministe deux cas sont possibles : ou bien la série résultante est formée de variables indépendantes, ou bien les variables semblent corrélées. Il peut être intéressant de prendre en compte dans le modèle cette corrélation. Il faut savoir modéliser une série de variables aléatoires dépendantes en nombre croissant. Pour cela, on fait appel à la notion de processus stochastique qui généralise les notions de variables et vecteurs aléatoires.

Définition 5. Soit T un ensemble d'indices infini ; $T = \mathbb{N}$ ou \mathbb{R} . Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On appelle processus stochastique sur T à valeur dans \mathbb{R} une application X mesurable qui à tout ω de Ω associe une fonction de T dans \mathbb{R} : $(X(t))_{t \in T}$.

Les fonctions $X(\omega)$ pour un ω fixé sont appelées les trajectoires du processus.

Définition 6. Un processus stochastique sur T est défini par la donnée d'une loi de probabilité sur \mathbb{R}^T , c'est-à-dire sur l'ensemble des fonctions $f : T \rightarrow \mathbb{R}$.

Remarque : la définition d'une telle loi de probabilité est généralement impossible explicitement. Cette loi n'est pas discrète (le nombre de fonctions est infini indénombrable) et devrait être définie par une densité par rapport à une mesure sur l'ensemble des fonctions \mathbb{R}^T . Mais il n'existe pas de mesure naturelle de l'ensemble \mathbb{R}^T comme la mesure de Lebesgue dans le cas de \mathbb{R} . La méthode pour définir une probabilité sur \mathbb{R}^T est la suivante. Il suffit de fixer une famille de probabilités (appelées lois marginales de dimension finie du processus) sur les vecteurs $(f(t_1), \dots, f(t_p))$ de toute taille p et pour tout (t_1, \dots, t_p) éléments de T en s'assurant que ces probabilités sont cohérentes entre elles (c'est-à-dire par exemple que la loi de la variable aléatoire $f(t_1)$ se déduit de la loi du couple $(f(t_1), f(t_2))$). Toute méthode qui permet de définir une telle famille de probabilités cohérentes définit bien un processus stochastique (théorème de Kolmogorov). Nous allons examiner plusieurs exemples :

7.6 Modèle de bruit blanc indépendant

Supposons que $T = \mathbb{N}$. On peut définir une famille de marginales de dimension finie par pour tout $t \in T$, $f(t)$ suit une loi G donnée et pour tout $s, t \in T$, si $s \neq t$, $f(s)$ et $f(t)$ sont deux variables indépendantes. Cette famille est cohérente.

7.7 Processus gaussiens généraux

Pour définir un processus gaussien sur T , il suffit de se donner une fonction espérance $m(t)$ et une fonction d'autocovariance $\Gamma(t_1, t_2)$ définie sur T^2 qui soit de type positif.

Définition 7. Une fonction $\Gamma(t_1, t_2)$ est de type positif si elle est symétrique et si pour tout entier p et p -uplet (t_1, \dots, t_p) de T et (u_1, \dots, u_p) de \mathbb{R} :

$$\sum_{i=1}^p \sum_{j=1}^p \Gamma(t_i, t_j) u_i u_j \geq 0.$$

Cette définition revient à dire que la matrice des $\Gamma(t_i, t_j)_{i=1, \dots, p, j=1, \dots, p}$ est toujours une matrice symétrique définie positive. On peut donc définir un vecteur gaussien $(X(t_1), \dots, X(t_p))$, ayant pour espérance $(m(t_1), \dots, m(t_p))$ et pour matrice de covariance cette matrice. On montre sans difficulté que cette définition donne des lois marginales de dimension finie cohérentes, en faisant appel aux propriétés particulières des vecteurs gaussiens.

7.8 Transformé instantané d'un processus

Si X est un processus sur T et si f est une fonction mesurable, alors Y définie par $Y_t = f(X_t)$ est un processus sur T . Il suffit de voir que les marginales finies sont bien définies par la relation qui lie X_t et Y_t et que les marginales de dimension finies restent cohérentes par la transformation par f ; il existe donc un processus Y qui correspond à ces marginales.

7.9 Modèles linéaires

Soit $(\varepsilon(t))_{t \in \mathbb{N}}$ un bruit blanc indépendant de loi \mathcal{L} de variance finie. Soit $(a_t)_{t \in \mathbb{N}}$ une suite de coefficients réels tels que $\sum_{s \in \mathbb{N}} a_s^2 < \infty$. On définit un processus X par

$$X(t) = \sum_{s \in \mathbb{N}} a_s \varepsilon(t + s).$$

Si les coefficients a_t sont nuls à partir d'un certain rang, cette définition est une généralisation simple du cas de la transformation instantanée. Si les coefficients a_t ne sont pas nuls à partir d'un certain rang, il faut montrer que la somme infinie a un sens dans l'espace des variables aléatoires ; en fait les variables $\varepsilon(t)$ font partie de l'espace normé complet des variables aléatoires de variance finie, noté $L^2(\mathbb{R})$; le fait que la somme des carrés des coefficients converge suffit à montrer que la série de variables aléatoires converge dans $L^2(\mathbb{R})$.

8 Modèle de prévision à court terme

On suppose que l'on dispose de données pour un phénomène jusqu'à la date d'aujourd'hui et on veut proposer une prévision de ce phénomène pour demain. La première idée consiste à regarder si les données ne peuvent pas se décomposer en une fonction déterministe du temps (tendance et saisonnalité) et un bruit blanc indépendant résiduel. Si ce processus est un processus indépendant, la valeur des résidus futurs est tout à fait imprévisible comme lorsqu'on lance un dé de façon successive ; la prévision la meilleure est alors donnée par la tendance déterministe. Si au contraire les résidus ne sont pas indépendants, il est possible de prévoir en partie les résidus futurs et donc de diminuer encore l'imprécision de la prévision.

8.1 Processus autorégressifs linéaires gaussiens

Nous allons définir une famille de processus sur \mathbb{N} permettant de décrire le processus des résidus. Ce modèle est adapté au problème de la prévision à court terme. Pour alléger les notations nous allons noter X_i la variable aléatoire correspondant à l'instant i de \mathbb{N} . Soit $(\varepsilon_i)_{i \in \mathbb{N}}$ un bruit blanc gaussien centré et de variance σ^2 . Soit X_0 une variable aléatoire gaussienne centrée de variance

σ_0^2 . Soit ρ un réel. Pour $i > 0$, on définit récursivement un processus gaussien centré autorégressif d'ordre 1 (AR1) par

$$X_{i+1} = \rho X_i + \varepsilon_{i+1}. \quad (3)$$

En substituant dans l'équation récursive (3), on montre que les composantes X_i du processus s'écrivent comme une combinaison linéaire des $(\varepsilon_j)_{j \leq i}$ et de X_0 . Le vecteur (X_0, \dots, X_i) est donc un vecteur gaussien et les marginales finies sont cohérentes. Il suffit pour identifier la loi du processus de calculer la fonction d'autocovariance $\Gamma_{i,j} = \text{cov}(X_i, X_j)$.

8.1.1 Processus stationnaire

Nous allons nous intéresser au cas où la série de données à modéliser (une fois la tendance déterministe retranchée) présente une certaine homogénéité au cours du temps. On fait l'hypothèse que la variance des variables aléatoires reste constante et que les relations de dépendance entre variables successives sont les mêmes au cours du temps.

Définition 8. Un processus est dit stationnaire si ses lois marginales sont invariantes par translation de temps. Pour tout (t_1, \dots, t_p) et tout k dans \mathbb{N} , les vecteurs $(X(t_1), \dots, X(t_p))$ et $(X(t_1 + k), \dots, X(t_p + k))$ ont la même loi.

Remarque : si un processus est gaussien de fonction de covariance $\Gamma(x, y)$ est stationnaire, on appelle fonction d'autocovariance la fonction $r(h) = r(x, x + h)$. Si le processus précédent défini par (3) est stationnaire, alors nécessairement X_0 et X_1 ont même variance. Or

$$\text{var}(X_1) = \rho^2 \text{var}(X_0) + \text{var}(\varepsilon_1) = \rho^2 \sigma_0^2 + \sigma^2.$$

Il faut donc que les paramètres définissant le modèle vérifient $(1 - \rho^2)\sigma_0^2 = \sigma^2$ ce qui impose que $|\rho| < 1$. Inversement, si cette relation entre les paramètres est vérifiée, on montre que $r(i, j) = r(i + k, j + k) = r(i - j)$ ce qui assure la stationnarité pour un processus gaussien.

8.1.2 Processus autorégressif d'ordre p

On peut généraliser le processus autorégressif d'ordre 1 à un processus d'ordre p ; on se donne ε le bruit blanc précédent, (X_0, \dots, X_{p-1}) un vecteur gaussien, (a_0, \dots, a_p) une suite de coefficients réels et on définit le modèle AR p par

$$a_0 X_i + a_1 X_{i-1} + \dots + a_p X_{i-p} = \varepsilon_i.$$

Le polynôme $P(z) = a_p z^p + a_{p-1} z^{p-1} + \dots + a_0$ est le polynôme associé au processus précédent. L'existence d'un processus stationnaire, obtenu pour un bon choix du vecteur initial (X_0, \dots, X_{p-1}) est assurée lorsque toutes les racines complexes ou réelles de ce polynôme sont de module strictement supérieur à 1.

8.1.3 Processus de moyenne mobile d'ordre q

On peut définir le processus de moyenne mobile d'ordre q (MA q) à partir du bruit blanc ε et de (b_0, \dots, b_q) une suite de coefficients réels par

$$X_i = b_0 \varepsilon_i + b_1 \varepsilon_{i-1} + \dots + b_q \varepsilon_{i-q}.$$

Le polynôme $Q(z) = b_0 z^q + b_1 z^{q-1} + \dots + b_q$ est le polynôme associé au processus précédent. Le processus MA q est toujours stationnaire par définition.

8.1.4 Processus ARMA(p, q)

On généralise les processus suivants par le processus ARMA(p, q) défini à partir du bruit blanc ε , d'un vecteur gaussien (X_0, \dots, X_{p-1}) , de deux suites (a_0, \dots, a_p) et (b_0, \dots, b_q) de coefficients réels par

$$a_0 X_i + a_1 X_{i-1} + \dots + a_p X_{i-p} = b_0 \varepsilon_i + b_1 \varepsilon_{i-1} + \dots + b_q \varepsilon_{i-q}.$$

Par convention on fixe $a_0 = b_0 = 1$. On considère les polynômes P et Q précédents et on les simplifie quand ils ont des racines communes. L'existence d'un processus stationnaire, obtenu pour un bon choix du vecteur initial (X_0, \dots, X_{p-1}) est assurée lorsque toutes les racines de P sont de module strictement supérieur à 1 et celles de Q de module supérieur à 1.

8.1.5 Processus intégré ARIMA

La série chronologique considérée peut, même après soustraction d'une tendance, apparaître comme variant au cours du temps de la façon suivante : l'espérance reste apparemment nulle, mais la variabilité s'accroît au cours du temps. La variance empirique des données calculée sur un intervalle augmente avec le temps. Un modèle simple ayant cette propriété est la marche aléatoire. Partant d'un bruit blanc ε centré de variance 1, on définit le processus de marche aléatoire

$$X_t = \sum_{i=1}^t \varepsilon_i.$$

Les variables X_t ne sont pas indépendantes et leur variance vaut t . Par contre, par définition, le processus des différences $X_t - X_{t-1}$ est un processus indépendant. Lorsque l'on observe une série dont la variance semble s'accroître, on peut calculer les différences discrètes de la série pour stabiliser la variance. Cette opération de différenciation peut être opérée plusieurs fois de suite. Les processus qui donnent des processus ARMA par différenciation forment la classe des processus ARIMA(p, d, q), où d est le nombre de fois qu'il faut différencier le processus pour obtenir un processus ARMA (p, q).

8.2 Prédiction par la méthode de Box et Jenkins

Soit une série chronologique de mesure X_1, \dots, X_n . On cherche à estimer la valeur prochaine de la série. On modélise la série par la somme d'une tendance et d'un processus ARMA(p, q). Les étapes de la prédiction sont les suivantes

- Calcul de la tendance $f(t)$ par la méthode de régression précédente.
- Modélisation de la série résultante U_t par un processus ARMA(p, q).
- Calcul de la prédiction et de son intervalle de confiance.

La première étape a été décrite précédemment. On cherche maintenant un modèle ARMA stationnaire approchant la série $U_t = X_t - f(t)$; il faut choisir les ordres p et q du modèle puis estimer les coefficients a_i et b_j ; il faut ensuite estimer les ε_t résidus correspondant et vérifier qu'ils forment bien un bruit blanc.

8.2.1 Choix des ordres (p, q)

Comme pour la régression multiple, il n'existe pas de meilleurs choix évidents ; plus les ordres sont grands, meilleure est l'adéquation, mais le nombre de paramètres à estimer augmente et leur estimation statistique devient imprécise. On emploie donc une règle heuristique en utilisant les corrélogrammes empiriques calculés à partir des données.

Définition 9. L'autocorrélogramme d'un processus stationnaire X est la série formée des $\rho_k = \text{corr}(X_i, X_{i+k})$.

Définition 10. L'autocorrélogramme partiel d'un processus stationnaire X est la série α_k telle que

- $\alpha_0 = 1$,
- α_1 est le coefficient de régression de X_i sur X_{i-1} ,
- α_2 est le coefficient de régression de X_i sur X_{i-2} , dans la régression sur l'espace engendré par $(X_{i-1}, X_{i-2}$ et ainsi de suite...

Propriété 4. Pour un processus AR_p , $\alpha_k = 0$ dès que $k > p$. Pour un processus MA_q , $\rho_k = 0$ dès que $k > q$.

On calcule l'autocorrélogramme empirique et l'autocorrélogramme partiel empirique de la série de données avec les intervalles de significativité ; on regarde les valeurs de k à partir desquelles les corrélogrammes ne sont plus significativement non nuls. Cela donne une borne raisonnable, généralement trop grande, pour le choix des p et q .

8.2.2 Estimation des paramètres

L'estimation des paramètres AR se fait par les équations de Yule Walker de façon explicite ; l'estimation des paramètres MA est plus difficile et demande un calcul itératif d'optimisation, et un certain nombre de fausses solutions doivent être écartées. L'ensemble de ces opérations est trop complexe pour être exposé ici. La plupart des logiciels traitant les séries chronologiques implémentent ces estimations sous le nom de méthode de Box et Jenkins.

8.2.3 Prévision

Connaissant les estimations des de la tendance f , des paramètres du modèle ARMA et les estimations du bruit ε , on propose d'estimer X_{n+1} par

$$\hat{X}_{n+1} = f(n+1) - \hat{a}_1(X_n - f(n)) - \cdots - \hat{a}_p(X_{n-p+1} - f(n-p+1)) + \hat{b}_1\hat{\varepsilon}_n + \cdots + \hat{b}_q\hat{\varepsilon}_{n-q+1}.$$

Le calcul de l'intervalle de confiance repose sur l'étude des covariances des différents estimateurs utilisés. Ces estimateurs étant très dépendants, le calcul est compliqué, mais il est généralement effectué par le logiciel qui calcule les estimations des paramètres. On peut proposer une prévision pour X_{n+2} de la même façon. On observe dans la formule la diminution du nombre de termes correspondant aux valeurs réellement observées. L'information apportée par le modèle ARMA diminue et la prévision se rapproche de la prévision déterministe définie par la fonction f . La prévision utilisant les modèles ARMA n'a d'intérêt qu'à court terme.

8.3 Conclusion sur la notion de dépendance

Le fait qu'une série d'observations ne soit pas indépendante a deux conséquences principales : Une conséquence négative : la dépendance entre les données modifie l'efficacité des techniques d'estimation, parfois de façon très lourde. Les méthodes d'estimation doivent être adaptées. Une conséquence positive : lorsque les variables d'une série sont dépendantes, les variables déjà observées ont un lien avec les valeurs futures; elles contiennent une information sur ces valeurs ce qui permet de construire de meilleures prévisions que dans des séries indépendantes. Sur le choix des modèles Un modèle stochastique doit posséder de nombreuses qualités. La première, souvent oubliée, est d'être correctement défini; deux cas de mauvaise définition peuvent se produire: ou bien le modèle est insuffisamment spécifié, c'est-à-dire que les probabilités de certains événements qui vont nous intéresser pour construire des méthodes statistiques ne peuvent être calculées à partir des éléments du modèle ; ou bien le modèle est défini par des conditions inappropriées; il n'existe aucune loi de probabilité vérifiant les conditions; le modélisateur peut néanmoins effectuer des calculs mathématiques exacts mais dénués de sens, pour parfois parvenir à une absurdité, comme une variance négative. La deuxième qualité est de correspondre aux données. Suivant le problème pratique, ce critère qui mesure cette adéquation va changer; il n'existe donc pas de bon modèle dans l'absolu. La troisième qualité, qui l'emporte souvent pour le choix du modèle, est d'être manipulable : il faut pouvoir calculer les lois de distributions de variables qui vont nous servir de statistiques, avoir des méthodes efficaces et facilement calculables d'estimation des paramètres. Lorsque cela n'est pas possible, on a recours à des méthodes de simulation. Un modèle dont on ne peut ni estimer les paramètres, ni simuler les trajectoires n'a pas d'utilité pratique. Ces trois qualités sont difficilement conciliables. L'art du modélisateur est de ne pas oublier la première et de ménager un compromis entre les deux autres.

8.4 Références

- Barbe P., Ledoux M. (1998) Probabilité, Collection Mathématiques : de la licence à l'agrégation, Belin, 252 p.
- Dacunha-Castelle, D., Duflo, M. (1983) Probabilités et statistiques. Tome 1 : Problèmes à temps fixe et Tome 2 : Problèmes à temps mobile. Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris, 286 p.