

STATISTIQUE des PROCESSUS. APPLICATIONS AUX FINANCES

PLAN

Avertissement au lecteur : ces notes de cours sont celles d'un cours de 20 heures que j'ai donné trois ans d'affilée au public mixte d'élèves de l'INSA et de SUPAERO. Il est essentiellement fondé sur les propres notes de DEA de Valentine GENON-CATALOT ; merci à elle de me les avoir communiquées, merci au temps qu'elle m'a consacré pour me rendre accessible ce domaine qui était complètement nouveau pour moi : elle n'a épargné ni son temps, ni sa patience ! Bien qu'elle ait relu une première version de ces notes, les (nombreuses !) fautes restantes sont uniquement de mon fait ! Merci à quiconque voudra bien me les signaler.

0.1 Introduction

- Objectifs : modélisation de phénomènes en continu : les cours des produits financiers, par des processus de diffusion observés à temps discrets ; estimation des paramètres de ces processus.
- Quelques modèles couramment utilisés (Black-Scholes, Vasicek, Cox-Ingersoll-Ross, modèle bilinéaire...)
- Construction du modèle, processus canonique, loi du processus.
- Rappels sur les équations différentielles stochastiques.
- Outils : vraisemblance, contraste.

0.2 Modèle de Black-Scholes

- Fonction de vraisemblance et estimateur associé du couple dérive-volatilité.
- Estimation de la volatilité par variation quadratique.
- Propriétés de ces estimateurs : biais, erreur quadratique.
- Asymptotique : lois limites, intervalles de confiance, consistance.

0.3 Modèle de Vasicek

- Existence d'une solution.
- Discrétisation du modèle pour obtenir un AR(1) ; estimateurs obtenus par les méthodes de Box et Jenkins
- Fonction de vraisemblance du modèle discret et estimateurs associés de θ et σ sur ce modèle.
- Estimateurs "naturels".
- Propriétés de ces estimateurs.
- Asymptotique.

0.4 Fonction de contraste sur le schéma d'Euler

- Définitions et généralités.
- Modèle de Brennan-Schwartz $dX_t = \alpha(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t$. Existence de solutions.
- Modèles du type $dX_t = \alpha(\beta - X_t)dt + \sigma X_t^\delta dW_t, \delta \in]\frac{1}{2}, 1]$. Existence de solutions.
- Estimateurs obtenus par une fonction de contraste sur les modèles discrétisés.
- Asymptotique.

Introduction, but du cours

Les objectifs de ce cours sont :

- . la modélisation de phénomènes en temps continu faisant intervenir le hasard, tels le cours des produits financiers (actifs, options, taux par exemple) par des processus stochastiques ;
- . l'estimation des paramètres de la dynamique de ces processus observés nécessairement à temps discret, et comme d'habitude lorsque l'on fait des estimations, la recherche de la vitesse de convergence, d'intervalles de confiance, etc.

Plus précisément, supposons que l'on dispose d'observations d'un processus $(X_t, t \in [0, T])$ qui évolue selon l'équation différentielle stochastique :

$$(1) \quad dX_t = b(t, X_t, \theta)dt + \sigma(t, X_t, \theta)dW_t ; t \in [0, T], X_0 = \eta,$$

sur un espace de probabilité filtré $(\Omega, \mathcal{A}, (\mathcal{F}_t), \mathbb{P})$, où W est un \mathcal{F} -mouvement brownien à valeurs dans \mathbb{R}^m , η une variable aléatoire \mathcal{F}_0 -mesurable (donc indépendante du mouvement brownien) de loi ν_θ . L'archétype d'un tel processus est le modèle de Black et Scholes :

$$dX_t = X_t b dt + X_t \sigma dW_t, X_0 = x,$$

et le paramètre θ est le couple (b, σ) tendance-volatilité dont la connaissance est cruciale pour l'évaluation des produits financiers.

Sinon, on suppose que l'on connaît les fonctions b et σ :

$$(2) \quad b : \mathbb{R}^+ \times \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d,$$

$$(3) \quad \sigma : \mathbb{R}^+ \times \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$$

ces fonctions ayant des propriétés de régularité suffisante pour que l'EDS ci-dessus ait une unique solution forte à θ fixé. (voir par exemple Protter [27].) On se pose le problème de l'estimation de θ à partir d'observations de la trajectoire dont on dispose. Les observations seront toujours de la forme $F_T(\xi_t, t \in [0, T])$. Notre but est donc de construire des estimateurs $\hat{\theta}$ de la "vraie" valeur θ_0 du paramètre, qui soient fonction des observations, puis d'en étudier les propriétés.

0.5 Différentes méthodes d'estimation

On obtient plusieurs sortes d'estimations avec des estimateurs de plusieurs types :

- estimateurs de **maximum de vraisemblance** (EMV), ce qui suppose de calculer la vraisemblance de la famille d'observations ;

- estimateurs de **minimum de contraste** : lorsque la vraisemblance de l'échantillon est difficile à calculer, voire impossible, ou difficile à exploiter, une alternative est de construire des pseudo-logvraisemblances dont l'opposé sera dit "contraste" et dont on tirera des estimateurs.

- estimateurs **empiriques** fondés sur une méthode de moments, éventuellement méthode de moments généralisée.

- estimateurs fondés sur des équations d'estimation, c'est à dire que si G_n est une fonction des observations (par exemple $G_n = \partial_\theta \log L$, gradient de la log-vraisemblance parfois appelée "fonction score"), une équation d'estimation est $G_n(\theta, x_i) = 0$ dont la solution suggère des estimations possibles de θ .

estimateur=statistique=variable aléatoire ; alors que estimation= valeur numérique prise par cette variable aléatoire au vu des observations.

0.6 Différents types d'observations

- observations discrètes : le point de vue naturel et le plus réaliste sur le plan pratique consiste à supposer que l'on observe la trajectoire ξ à des instants t_1, \dots, t_N , avec $0 \leq t_1 \leq \dots \leq t_N \leq T$. Le modèle statistique est alors défini par l'observation du N -uplet $(\xi_{t_1}, \dots, \xi_{t_N})$. On étudiera notamment le cas où les observations sont faites régulièrement : $t_i = i\Delta, T = N\Delta$, Δ est le pas de la discrétisation. Eventuellement, le pas Δ_N dépend du nombre d'observations.

- observation continue: on suppose que l'on observe continuellement la trajectoire ξ sur tout l'intervalle de temps $[0, T]$. Le modèle statistique est alors défini par l'observation de la fonction $\xi : [0, T] \rightarrow E$, et les estimateurs de θ seront des fonctions de cette trajectoire.

Cette hypothèse est admise dans toute la littérature portant sur la statistique des processus à temps continu. Elle permet les études asymptotiques lorsque l'on fait les observations à des instants de plus en plus rapprochés. Elle peut aussi être justifiée dans le domaine de la finance où l'on dispose de données assez rapprochées dans le temps (cotation en continu d'indices boursiers, par exemple). Mais cette justification est sujette à caution ; en effet, quelques études empiriques tendent à faire penser que les phénomènes sont différents selon la fréquence d'observation, comme s'il y avait un modèle en basse fréquence et un autre en haute fréquence. Voir par exemple les travaux de J.M. Bardet.

0.7 Différents cadres asymptotiques

L'étude et la comparaison des différents estimateurs n'est possible, dans la plupart des cas, que dans un cadre asymptotique. En statistique des diffusions, plusieurs approches asymptotiques sont envisageables, qui diffèrent suivant les modèles que l'on considère ou suivant les observations dont on dispose.

- On pourra faire tendre T vers l'infini (longueur de l'intervalle d'observation), ou bien N vers l'infini (nombre d'observations), Δ restant fixe.

- On peut également supposer simultanément que N tend vers l'infini et Δ tend vers 0, le produit T étant fixe (observations de plus en plus rapprochées à l'intérieur d'un intervalle fixe) : par exemple $\Delta_N = \frac{T}{N}$, $N \rightarrow \infty$, $\Delta_N \rightarrow 0$.

- On peut enfin supposer que N et T tendent vers l'infini, même si Δ_N tend vers 0 : observations de plus en plus rapprochées à l'intérieur d'un intervalle de plus en plus grand : par exemple $\Delta_N = \frac{a}{\sqrt{N}}$, $T_N = a\sqrt{N}$.

Selon le cas, on identifiera les paramètres inconnus figurant dans les fonctions b et/ou σ .

- Enfin, soit $\varepsilon > 0$, et supposons que le coefficient de diffusion soit de la forme $\varepsilon\sigma(t, x)$, σ étant complètement connu. Le modèle dans ce cas est celui d'une "petite perturbation" de l'équation différentielle ordinaire

$$dx(t) = b(t, x(t), \theta)dt,$$

$$dx_t^\varepsilon = b(t, x_t^\varepsilon, \theta)dt + \varepsilon\sigma(t, x_t^\varepsilon)dW_t,$$

Le cadre asymptotique d'étude est alors la recherche de la limite lorsque ε tend vers 0.

Si l'on se réfère aux modèles usuels, une situation asymptotique est une situation où l'on fait de plus en plus d'observations jusqu'à atteindre une infinité d'observations. C'est le cas cité plus haut lorsque la longueur T de l'intervalle tend vers l'infini ou lorsque le pas Δ de discrétisation tend vers 0.

C'est le cas aussi de la petite perturbation – qui correspond à un modèle à petite variance – comme le montre l'exemple particulier suivant.

Exemple : soient n processus réels (Y_i) indépendants et de même loi, dont la dynamique est donnée par

$$dY_t^i = [a(t, \theta)Y_t^i + b(t, \theta)]dt + dW_t^i ; t \in [0, T], Y_0^i = y \in \mathbb{R},$$

où W^i sont n mouvements browniens indépendants. On a ainsi une dérive linéaire $b(t, x, \theta) = a(t, \theta)x + b(t, \theta)$, et un coefficient de diffusion constant égal à 1. Si l'on considère maintenant le processus moyen $\bar{Y}_t = (Y_t^1 + \dots + Y_t^n)/n$, celui-ci est de dynamique

$$d\bar{Y}_t = [a(t, \theta)\bar{Y}_t + b(t, \theta)]dt + \frac{1}{\sqrt{n}}dB_t^n ; t \in [0, T], \bar{Y}_0 = y \in \mathbb{R},$$

où $B^n = \frac{W^1 + \dots + W^n}{\sqrt{n}}$ est un mouvement brownien. Ainsi, \bar{Y} est régi par une EDS de petite variance $\frac{1}{\sqrt{n}}$ quand n tend vers l'infini. **Le cas de petite variance s'apparente au cas où l'observation est une moyenne d'observations indépendantes et de même loi de diffusion.**

1 Exemples de modèles

1.1 Actifs financiers, “stocks”

Considérons un marché financier modélisé constitué de d actifs dont on considère que leur dynamique est une diffusion sur un espace de probabilité filtré $(\Omega, \mathcal{A}, (\mathcal{F}_t), \mathbb{P})$, la filtration étant la filtration naturelle engendrée par un n -mouvement brownien W . On note S_1, \dots, S_d les processus de prix des actifs :

$$dS_t^i = S_t^i(b^i(t, S_t)dt + \sigma_j^i(t, S_t)dW_t^j), i = 1, \dots, d,$$

où les fonctions b et σ peuvent dépendre de paramètres à estimer. Certaines actions peuvent ne pas être observables.

1.2 Taux d'intérêt

Pour $h > 0$, le taux d'intérêt R_t^h observé à la date t et de terme h , $(1 + R_t^h = \frac{P_{t+h}}{P_t})$ fait aussi l'objet de modélisations stochastiques, que l'on appelle “modèles de courbes de taux”. Supposant que la limite de R_t^h existe quand h tend vers 0, notée r_t , elle définit un processus appelé “taux cible” ou “spot”). Les modèles suivants ont été proposés pour ce taux :

$$dr_t = \alpha(\nu - r_t)dt + \rho dW_t, \text{ (Musiela)}$$

ou :

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t \text{ (Heath-Jarrow-Morton)}.$$

Il en existe bien d'autres ; en particulier, σ peut être fonction du temps.

Notons bien sûr que les modélisations par des processus de diffusion, solutions d'EDS, existent dans bien d'autres domaines que celui de la finance : génétique, mécanique, météorologique par exemple.

1.3 Modèles couramment utilisés en finance

Le premier modèle est celui présenté par Bachelier dans sa thèse en 1900 [1]. Il s'agissait d'un modèle gaussien :

$$S_t = bt + \sigma W_t, t \in [0, T], W \text{ mouvement brownien standard.}$$

Les modèles ont été ensuite plus sophistiqués, mais celui-ci, à l'époque, s'ajustait assez bien aux données de la Bourse de Paris.

1.3.1 Black-Scholes (1973)

De fait le problème traité par Black et Scholes dans [5] est celui de l'évaluation et la couverture d'une option de type européen sur une action de dividende. Ce modèle est très utilisé, d'où l'importance de bien estimer les paramètres du modèle : les fonctions b et σ ci-dessus sont simplement des constantes, donc ce sont les deux paramètres à estimer ; il n'y a qu'une action :

$$dS_t = S_t(bdt + \sigma dW_t), t \in [0, T], W \text{ mouvement brownien standard.}$$

L'interprétation de cette EDS est la suivante : $d \log S_t \sim \frac{S_{t+h} - S_t}{S_t} \sim bh + \sigma(W_{t+h} - W_t)$, est l'accroissement relatif du prix de l'actif, soit son rendement sur une petite durée h .

On dit que la loi est "log-normale" (le log de l'actif suit une loi gaussienne, dont un petit accroissement représente le rendement de l'action. Il s'agit d'une EDS linéaire explicitement résoluble : $S_t = \exp(\sigma W_t + (b - \frac{1}{2}\sigma^2)t)$. On pourrait même prendre $S_t = \exp(\sigma W_t + bt)$. En effet, dans la pratique, on trouve comme ordres de grandeur $b \sim 10^{-3}$, $\sigma^2 \sim 10^{-7}$ en log de Francs par minute.

Définition 1.1 b s'appelle la tendance, σ la volatilité.

On peut également prendre d actifs, S sera à valeurs dans \mathbb{R}^d , b un vecteur et σ une matrice. Ce modèle sera étudié en détail dans le chapitre 2.

1.3.2 Vasicek

ou plutôt, comme l'on dit en calcul stochastique, processus de Ornstein-Uhlenbeck :

$$dS_t = \mu S_t dt + \sigma dW_t, t \in [0, T], S_0 = \eta,$$

ou bien :

$$dS_t = \alpha(\nu - S_t)dt + \sigma dW_t, t \in [0, T], S_0 = \eta,$$

η variable aléatoire gaussienne de moyenne m et variance τ^2 , indépendante de W . Le paramètre à estimer est $\theta = (\mu, \sigma)$. Voir [9] page 191 où la solution explicite est donnée :

$$S_t = (\eta - \nu)e^{-\alpha t} + \nu + \sigma \int_0^t e^{-\alpha(t-u)} dW_u.$$

Exercice : retrouver par le calcul de Itô l'EDS à partir de l'expression explicite de S_t .

1.3.3 Cox-Ingersol-Ross

C'est un modèle utilisé principalement pour la courbe des taux ([9] page 188) :

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t.$$

Hull et White :

$$dr_t = (\theta + a_t(b - r_t))dt + \sigma_t dW_t,$$

1.3.4 Modèle bilinéaire

$$dX_t = (aX_t + b)dt + \sigma X_t dW_t.$$

1.3.5 Diffusion élastique

$$dX_t = (aX_t + b)dt + \sigma X_t^\psi dW_t, \psi \in]0, 1[.$$

1.4 Construction du modèle -processus canonique-loi du processus

Rappelons le modèle général donné par la dynamique (1). Il s'agit d'une diffusion (cf cours M. G. GAUDRON)

Définition 1.2 Une diffusion est le processus stochastique solution d'une telle EDS.

Remarquons qu'à l'origine le terme de diffusion désigne un processus de Markov fort à trajectoires continues (cf Karlin-Taylor, 1981, chapitre 15).

1.4.1 Le modèle canonique (théorème de Kolmogorov)

C'est l'espace de probabilité filtré $(\Omega, \mathcal{A}, (\mathcal{F}_t), \mathbb{P})$ où :

- Ω est identifié à l'espace des trajectoires continues de $[0, T]$ dans \mathbb{R} , noté C_T , muni de la topologie de TCU : $d(x, y) = \sup_{0 \leq t \leq T} |x(t) - y(t)|$. On note \mathcal{B} sa tribu borélienne pour la topologie TCU.

- le processus canonique est $X : \Omega \times [0, T] \rightarrow \mathbb{R}, (\omega, t) \mapsto \omega(t)$.

- on note \mathcal{C}_T la plus petite tribu qui rend mesurables les applications coordonnées:

$$X_t : \Omega \rightarrow \mathbb{R},$$

$$\omega \mapsto \omega(t), X_t(\omega) = \omega(t).$$

Les applications coordonnées sont continues pour la TCU : $|\omega_1(t) - \omega_2(t)| \leq d(\omega_1, \omega_2)$.

On peut montrer de façon plus précise :

Théorème 1.3 La tribu borélienne et \mathcal{C}_T coïncident avec la tribu cylindrique $\mathcal{A} = \sigma\{\mathcal{U}\}$ où \mathcal{U} est la classe des cylindres, soit

$$\mathcal{U} = \{A = \{x \in \Omega, (x_{t_1}, \dots, x_{t_k}) \in B, B \in \mathcal{B}(\mathbb{R}^k), 0 \leq t_1 \leq \dots \leq t_k \leq T\}.$$

Preuve

(i) Tout élément $A \in \mathcal{U}$ se décrit avec des images réciproques d'applications coordonnées. Cette remarque montre l'inclusion $\sigma(\mathcal{U}) \subset \sigma(X_s, s \in [0, T])$.

(ii) Pour tout $s \in [0, T]$, l'application X_s est continue pour la TCU, donc \mathcal{B} -mesurable et l'on

a l'inclusion $\mathcal{C}_T \subset \mathcal{B}$.

(iii) On utilise le fait que la topologie TCU sur Ω est séparable, ce qui entraîne que la tribu des boréliens \mathcal{B} est engendrée par la classe des boules de la forme

$$B(a, r) = \{x \in \Omega; \sup_{0 \leq t \leq T} |x_t - a_t| \leq r\}$$

où $a \in \Omega, r \geq 0$. Or la continuité des éléments de $\Omega = C_T$ permet d'écrire

$$B(a, r) = \{x \in \Omega; \sup_{t \in [0, T] \cap \mathbb{Q}} |x_t - a_t| \leq r\} = \bigcap_{t \in [0, T] \cap \mathbb{Q}} \{x \in C_T; |x_t - a_t| \leq r\}$$

et cette intersection dénombrable est élément de \mathcal{A} ce qui montre l'inclusion $\mathcal{B} \subset \mathcal{A}$. \square

- filtration canonique $(\mathcal{F}_t, t \in [0, T])$ où la filtration est continue à droite :

$$\mathcal{F}_t = \bigcap_{s > t} \mathcal{G}_s, \mathcal{G}_s = \sigma(\mathcal{U}_{[0, s]}) = \sigma(X_u, u \leq s).$$

Remarque : par construction, le processus X est \mathcal{F} -adapté.

1.4.2 Loi du processus

On construit la probabilité \mathbb{P} sur (Ω, \mathcal{A}) (VGC pages (5-6-7)), c'est à dire que l'on doit construire cette probabilité en sorte qu'elle donne la loi de la variable aléatoire X qui est à valeurs dans l'espace des trajectoires C_T muni de la tribu cylindrique.

Définition 1.4 La loi de probabilité \mathbb{P}^T sur (C_T, \mathcal{C}_T) image de \mathbb{P} par la variable aléatoire X s'appelle la loi de probabilité du processus $(X_t, t \in [0, T])$. Elle est définie par

$$\mathbb{P}^T(B) = \mathbb{P}\{\omega : X(\omega) \in B\}, B \in \mathcal{C}_T.$$

Exemples :

(i) $B = \{x \in C_T, x(t_1) \leq a_1, \dots, x(t_n) \leq a_n\}$ où $t_1 \leq \dots \leq t_n, a \in \mathbb{R}^n$. Alors $\mathbb{P}^T(B) = \mathbb{P}\{\omega : \omega(t_1) \leq a_1, \dots, \omega(t_n) \leq a_n\}$.

(ii) $B = \{x \in C_T, \sup_{t \in [0, T]} x(t) > a\}$. Dans ce cas, $\mathbb{P}^T(B) = \mathbb{P}\{\omega : \sup_{t \in [0, T]} X_t(\omega) > a\}$.

(iii) Plus généralement, si f est une application mesurable sur (C_T, \mathcal{C}_T) ,

$$E_{\mathbb{P}}[f] = \int_{\Omega} f \circ X(\omega) d\mathbb{P}(\omega) = \int_{C_T} f(x) d\mathbb{P}^T(x) = E_{\mathbb{P}^T}[f].$$

Par exemple, si X est le mouvement brownien standard, \mathbb{P} est définie sur (Ω, \mathcal{A}) en sorte que l'image de \mathbb{P} par le vecteur aléatoire $(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$ est la loi gaussienne $\mathcal{N}(0, t_1) \otimes \dots \otimes \mathcal{N}(0, t_n - t_{n-1})$.

De façon générale, la probabilité \mathbb{P} est définie sur (Ω, \mathcal{A}) en sorte que l'image de \mathbb{P} par une application cylindrique est la loi d'un vecteur discrétisé du processus.

Proposition 1.5 (VGC page 7) La loi de probabilité \mathbb{P} est caractérisée par ses répartitions finies, c'est à dire : il suffit de donner $\forall k \in \mathbb{N}, \forall t_1 \leq \dots \leq t_k, \forall a_1, \dots, a_k \in \mathbb{R}^k$, la quantité $\mathbb{P}\{x, x_{t_i} \leq a_i, i = 1, \dots, k\}$.

Cette proposition justifie l'intérêt de l'introduction des cylindres.

Preuve : ces données définissent \mathbb{P} sur la classe \mathcal{U} qui est une classe monotone, puis on utilise le théorème de Carathéodory pour conclure. \square

Premier exemple : reprenons le mouvement brownien standard, \mathbb{P} est définie sur (Ω, \mathcal{A}) en sorte que l'image de \mathbb{P} par le vecteur aléatoire $(X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$ est la loi gaussienne $\mathcal{N}(0, t_1) \otimes \dots \otimes \mathcal{N}(0, t_n - t_{n-1})$, c'est à dire que

$$\mathbb{P}(X_0 = 0) = 1;$$

$$\mathbb{P}(X_{t_i} - X_{t_{i-1}} \leq a_i, i = 1, \dots, k) = \prod_1^k \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{-\infty}^{a_i} e^{-\frac{u^2}{2(t_i - t_{i-1})}} du.$$

Deuxième exemple (VGC pages 8-9) :

$$X_t = W_t + \int_0^t f_s^\theta ds, \text{ où } f \in L_{loc}^1(\mathbb{R}^+, \mathbb{R}).$$

Le vecteur $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ est sous \mathbb{P} de loi $F(\otimes_{i=1}^n \mathcal{N}(\int_{t_{i-1}}^{t_i} f_s^\theta ds, t_i - t_{i-1}))$ où F est l'application de \mathbb{R}^n dans $\mathbb{R}^n : x \mapsto (y_i = x_1 + \dots + x_i)$. Si $0 = t_0 < t_1 < \dots < t_k \leq T$, $(X_{t_i} - X_{t_{i-1}}, i = 1, \dots, k)$ est un vecteur aléatoire de composantes indépendantes et chacune de loi gaussienne de moyenne $\int_{t_{i-1}}^{t_i} f^\theta(s) ds$ et de variance $t_i - t_{i-1}$. On peut comme dans l'exemple précédent construire la probabilité \mathbb{P}^T sur C_T .

Exercice : écrire la loi dans ce cas.

La densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n est

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp - \frac{(y_i - y_{i-1} - \int_{t_{i-1}}^{t_i} f_s^\theta ds)^2}{2(t_i - t_{i-1})}$$

obtenue par loi image par l'application F (dont le jacobien est 1) de

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp - \frac{(x_i - \int_{t_{i-1}}^{t_i} f_s^\theta ds)^2}{2(t_i - t_{i-1})}.$$

1.4.3 Modèle statistique associé

Lorsque la loi du processus dépend d'un paramètre θ à déterminer (estimer), le modèle est l'ensemble $(\Omega, \mathcal{A}, (\mathcal{F}_t, t \in [0, T]), (\mathbb{P}^\theta, \theta \in \Theta))$. Ainsi, dans l'exemple ci-dessus, \mathbb{P}^θ est la probabilité sur $\Omega = C_T$ muni de ses boréliens telle que $F[\otimes_{i=1}^n \mathcal{N}(\int_{t_{i-1}}^{t_i} f(s, \theta) ds, t_i - t_{i-1})]$ est la loi sous \mathbb{P}^θ d'un échantillon du processus $(X_t^\theta, t \in [0, T])$, θ étant le paramètre à estimer dans Θ .

1.5 Rappels sur les EDS

(cours de G. GAUDRON et R. CONT)

- définition,
- hypothèses suffisantes d'unicité, d'existence,
- théorème de Girsanov(Cameron-Martin), lien avec les probabilités "risque neutre",
- formule de Itô,
- Feynman-Kac.
- martingales.

1.6 Outils qui seront utilisés

1.6.1 Vraisemblance, log-vraisemblance

Il s'agit d'outils d'usage facile, avec des hypothèses faibles, dans le cas où $\forall \theta \in \Theta$ il existe une probabilité \mathbb{P} qui domine toutes les probabilités \mathbb{P}^θ .

Définition 1.6 On appelle vraisemblance de θ la fonction sur $\Theta \rightarrow \mathbb{R}^+$, $\theta \mapsto \frac{d\mathbb{P}^\theta}{d\mathbb{P}}(x)$, c'est à dire la densité prise pour l'observation x .

Cette fonction est d'une certaine mesure liée au théorème de Cameron-Martin (cf cours de G. GAUDRON) ; en effet, dans le dernier exemple où $dX_t^\theta = f_t^\theta dt + dW_t$, il est connu que X^θ suit une loi de mouvement brownien sous la probabilité équivalente $Q = L\mathbb{P}$,

$$L = \exp\left[-\int_0^T f_t^\theta dW_t - \frac{1}{2} \int_0^T (f_t^\theta)^2 dt\right].$$

On cherche à définir \mathbb{P}^θ , loi de X^θ sous \mathbb{P} . Pour toute fonction g borélienne bornée sur C_T , d'une part

$$E_Q[g(X^\theta)] = E_{\mathbb{P}}[g(W)],$$

et d'autre part

$$E_Q[g(X^\theta)] = E_{\mathbb{P}}[Lg(X^\theta)],$$

Et si l'on utilise que $\mathbb{P} = L^{-1}.Q$ et $dW_s = dX_s^\theta - f_s^\theta ds$ et $E_{\mathbb{P}^\theta}[g] = E_{\mathbb{P}}[g(X^\theta)]$:

$$\begin{aligned} E_{\mathbb{P}}[g(X^\theta)] &= E_Q[\exp(\int_0^T f_s^\theta dW_s + \frac{1}{2} \int_0^T (f_s^\theta)^2 ds)g(X^\theta)] \\ &= E_Q[\exp(\int_0^T f_s^\theta dX_s^\theta - \frac{1}{2} \int_0^T (f_s^\theta)^2 ds)g(X^\theta)] \\ &= E_{\mathbb{P}}[\exp(\int_0^T f_s^\theta dW_s - \frac{1}{2} \int_0^T (f_s^\theta)^2 ds)g(W)] \end{aligned}$$

montre que $\mathbb{P}^\theta = \exp(\int_0^T f_s^\theta dW_s - \frac{1}{2} \int_0^T (f_s^\theta)^2 ds) \cdot \mathbb{P}^W$, soit :

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}^W}(x) = \exp[\int_0^T f_t^\theta dx_t - \frac{1}{2} \int_0^T (f_t^\theta)^2 dt],$$

et la fonction de vraisemblance

$$\theta \mapsto \exp[\int_0^T f_t^\theta dx_t - \frac{1}{2} \int_0^T (f_t^\theta)^2 dt]$$

d'où l'introduction de la log-vraisemblance,

$$\theta \mapsto \int_0^T f_t^\theta dx_t - \frac{1}{2} \int_0^T (f_t^\theta)^2 dt$$

qui se justifie également dans tout modèle de type exponentiel.

Définition 1.7 On appelle log-vraisemblance de θ la fonction sur $\Theta \rightarrow \mathbb{R}^+$, $\theta \mapsto \log \frac{d\mathbb{P}^\theta}{d\mathbb{P}}(x)$.

Exercice : Dans un modèle où

$$dX_t^\theta = f_t^\theta dt + \sigma_t dW_t, \sigma_t \neq 0, dt \otimes d\mathbb{P} \text{ presque sûrement, } X_0^\theta = x_0,$$

(VGC 2.2, pages 16-18 bis), \mathbb{P}^θ est la loi de X^θ , \mathbb{P} celle de $M_t = x_0 + \int_0^t \sigma_s dW_s$. Par le théorème de Cameron-Martin, on obtient

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}}(x) = \exp[\int_0^T \sigma_t^{-2} f_t^\theta dx_t - \frac{1}{2} \int_0^T (\sigma_t^{-1} f_t^\theta)^2 dt],$$

et sous \mathbb{P}^θ , $\int_0^t \sigma_s^{-1} (dX_s^\theta - f(\theta, s) ds)$ suit une loi de mouvement brownien, d'où pour ce modèle la log-vraisemblance :

$$\int_0^T \sigma_t^{-2} f_t^\theta dx_t - \frac{1}{2} \int_0^T (\sigma_t^{-1} f_t^\theta)^2 ds.$$

Exemple :

$$(4) \quad f(\theta, s) = \theta, \sigma_s = \sigma \sqrt{1+s^2},$$

σ étant connu. Ici la log-vraisemblance est

$$\frac{\theta}{\sigma^2} \int_0^T \frac{dx_s}{1+s^2} - \frac{1}{2} \frac{\theta^2}{\sigma^2} \text{Arctg}T$$

qui est une fonction concave en θ et atteint son maximum pour $\hat{\theta} = \frac{1}{\text{Arctg}T} \int_0^T \frac{dx_s}{1+s^2}$.

Comme $\frac{dx_s}{1+s^2} = \frac{\theta ds}{1+s^2} + \frac{\sigma dB_s}{\sqrt{1+s^2}}$, $\hat{\theta}$ est de loi gaussienne $\mathcal{N}(\theta, \frac{\sigma^2}{\text{Arctg}T})$. C'est un estimateur sans biais, mais n'a aucune propriété de convergence lorsque T tend vers l'infini : ainsi, l'erreur quadratique d'estimation décroît vers $2\sigma^2/T$ sans tendre vers 0.

1.6.2 Loi des estimateurs de maximum de vraisemblance, comportement asymptotique

(VGC, 2.3, pages 18bis-19ter)

Considérons l'exemple particulier du modèle

$$dX_t = \theta f(t)dt + \sigma(t)dB_t, x_0 = 0,$$

où f et σ sont des fonctions connues et déterministes comme dans les exemples précédents. La fonction de log-vraisemblance s'écrit

$$l_T(\theta) = \theta \int_0^T \frac{f(s)}{\sigma^2(s)} dx_s - \frac{\theta^2}{2} \int_0^T \frac{f^2(s)}{\sigma^2(s)} ds,$$

concave en θ , d'où l'on tire l'EMV

$$\hat{\theta}_T = \frac{\int_0^T \frac{f(s)}{\sigma^2(s)} dx_s}{\int_0^T \frac{f^2(s)}{\sigma^2(s)} ds}.$$

Dans cet exemple, la loi de $\hat{\theta}$ est facile à trouver : c'est une gaussienne puisque $dX_t = \theta f_t dt + \sigma_t dB_t$, et $\hat{\theta} = \frac{\int_0^T \sigma_s^{-2} f_s (\theta f_s ds + \sigma_s dB_s)}{\int_0^T \sigma_s^{-2} f_s^2 ds}$ est de loi $\mathcal{N}(\theta, (\int_0^T \sigma_s^{-2} f_s^2 ds)^{-1})$ car $\int_0^T \sigma_s^{-2} f_s dX_s = \theta \int_0^T \sigma_s^{-2} f_s^2 ds + \int_0^T \sigma_s^{-1} f_s dB_s$.

L'estimateur de maximum de vraisemblance $\hat{\theta}$ est donc sans biais et son comportement asymptotique dépend de $\int_0^T \sigma_s^{-2} f_s^2 ds$. Il n'y a pas de théorie générale ; tout dépend de ce que sont σ et f : il y a convergence L^2 (donc en probabilité) dès que $\sigma_s^{-2} f_s^2$ n'est pas intégrable sur \mathbb{R}^+ , convergence presque sûre chaque fois que l'on peut mettre l'estimateur sous la forme d'une moyenne de variables aléatoires intégrables indépendantes et de même loi (loi des grands nombre).

On a plus généralement la convergence presque sûre par un théorème de convergence de martingales : si M est une martingale locale continue, telle que $\langle M \rangle_t$ converge presque sûrement vers l'infini quand t tend vers l'infini, alors

$$\frac{M_t}{\langle M \rangle_t} \rightarrow 0.$$

Dans cet exemple, l'estimateur est bien une martingale et on peut conclure ici que $\hat{\theta}_T$ converge presque sûrement vers θ à l'infini dès que $\langle M \rangle_t = (\int_0^t \sigma_s^{-2} f_s^2 ds)^{-1}$ converge quand t tend vers l'infini.

Dans l'exemple (4), l'estimateur $\hat{\theta}$ n'est même pas consistant puisque $\text{Arctg } T$ converge vers $\pi/2$ pour T infini, et l'erreur quadratique d'estimation décroît vers $2\sigma^2/T$ sans tendre vers 0.

On peut finalement retenir quelques faits. Ou bien les EMV se calculent de manière explicite en fonction des processus observés (c'est le cas des précédents exemples) et on peut alors étudier directement ces estimateurs.

Sinon, on les étudie directement sur l'équation qui les définit. La forme explicite des fonctions de dérive $b(t, \theta)$ et de volatilité $\sigma(t, \theta)$ fournit les éléments de l'étude asymptotique des EMV lorsque T tend vers l'infini.

Exemple : $dX_t = f(\theta t)dt + dB_t, X_0 = 0$, (cf [19])

où f est de période 1 sur \mathbb{R} , de classe $C^\infty, \theta \in \mathbb{R}$. Dans cet exemple, la log-vraisemblance est

$$l(X, \theta) = \int_0^T f(\theta t)dX_t - \frac{1}{2} \int_0^T f^2(\theta t)dt,$$

dont on étudie le comportement avec le calcul des deux premières dérivées

Exercice : on pose $I = \frac{1}{3} \int_0^1 (f'(u))^2 du = \lim_{T \rightarrow \infty} T^{-3} \int_0^T t^2 (f'(\theta t))^2 dt$.

$$\frac{\partial l}{\partial \theta} = \int_0^T t f'(\theta t)dX_t - \int_0^T t f'(\theta t)f(\theta t)dt, \quad \frac{\partial^2 l}{\partial \theta^2} = \int_0^T t^2 f''(\theta t)dX_t - \int_0^T t^2 [(f'^2 + f''f)(\theta t)]dt,$$

(VGC page 19 ter, et [19] page 209) Montrer que $\frac{1}{T^3} l''(\theta)$ converge en probabilité vers $-I$, et $\frac{1}{T^{3/2}} l'(\theta)$ converge en loi vers une loi normale centrée de variance I . Alors si $\hat{\theta}$ réalise le maximum de la log-vraisemblance, on peut montrer la convergence en loi :

$$T^{3/2}(\hat{\theta} - \theta) \rightarrow \mathcal{N}(0, I^{-1}).$$

(cf théorème de Slutsky).

1.6.3 Contraste

Quand on ne peut rien tirer de la vraisemblance, une alternative est de considérer le schéma d'Euler sur les observations discrétisées, par exemple :

$$X_{k\Delta} - X_{(k-1)\Delta} \sim \Delta b(\theta, X_{(k-1)\Delta}) + \sigma[W_{k\Delta} - W_{(k-1)\Delta}].$$

Noter qu'ici, lorsque b n'est pas linéaire, le modèle n'est plus gaussien.

Ce processus approché est une chaîne de Markov de loi de transition

$$p(x, y; \theta) = \mathcal{N}(x + \Delta b(\theta, x), \sigma^2 \Delta)(y).$$

Définition 1.8 On appelle *contraste* (ou *fonction de contraste*) l'application sur Θ qui à θ fait correspondre, par exemple, l'opposé du logarithme de la densité de la chaîne de Markov définie par le schéma d'Euler prise aux observation discrètes du processus étudié, et $\tilde{\theta}_\Delta$ qui minimise ce contraste s'appelle l'estimateur de minimum de contraste (EMC).

Ceci donne pour l'exemple précédent lorsque $\sigma = 1$,

$$X_{(k+1)\Delta} - X_{k\Delta} \sim f(\theta k\Delta)\Delta + W_{(k+1)\Delta} - W_{k\Delta},$$

la probabilité de transition $p_t(x, y; \theta)$ est la densité d'une gaussienne de moyenne $x + \Delta f(\theta t)$ et de variance Δ . La pseudo vraisemblance pour des observations (x_i) aux temps $i\Delta$ est

$$\prod_i p(x_i, x_{i+1}; \theta) = \exp - \sum_i \frac{[x_{i+1} - x_i - \Delta f(\theta i\Delta)]^2}{2\Delta}$$

ce qui donne comme contraste

$$C(\theta) = \sum_i (x_{i+1} - x_i - \Delta f(\theta i\Delta))^2.$$

Exercice : chercher le minimum de contraste.

$$\frac{1}{2\Delta^2} C'(\theta) = \sum_i i f'(\theta i\Delta) (x_i + \Delta f(\theta i\Delta) - x_{i+1})$$

$$\frac{1}{2\Delta^2} C''(\theta) = \sum_i i^2 \Delta f''(\theta i\Delta) (x_i + \Delta f(\theta i\Delta) - x_{i+1}) + (i\Delta)^2 (f'(\theta i\Delta))^2.$$

2 Modèle de Black et Scholes

“Louis Bachelier a soutenu sa thèse le 29 Mars 1900. Son oeuvre prophétique n’a d’abord eu aucune reconnaissance de la part des mathématiciens français tels Borel et Lévy par exemple. En revanche, elle a été d’abord reconnue à Moscou par Khinchine et Kolmogorov vers 1930, avant d’être redécouverte in extremis par Lévy en 1944 (Bachelier est mort en 1946 !) et plus tard par la communauté internationale dans les années soixante[...]

Borel, pour sa part, se contente de considérer Bachelier comme un mathématicien assez faible (il ne fait partie du sérail à aucun titre, ni normalien, ni polytechnicien, une mention “honorable” seulement pour sa thèse de mathématiques, autant dire rien du tout) mais original et en tout cas méritant, qu’il convient d’aider dans une certaine mesure si l’occasion s’en présente [...]

Signalons qu’en revanche Bachelier s’est vu refuser par la commission de mathématiques de la faculté des sciences de Paris en 1909 une subvention spéciale pour l’impression de son livre (qui sera publié à compte d’auteur en 1912) avec ce rapport de Jules Tannery : “ M. Bachelier a demandé une subvention de 3000 francs pour lui permettre d’imprimer ses recherches sur le calcul des probabilités. La commission ne propose pas d’accorder cette subvention, d’autant que les travaux que M. Bachelier voudrait éditer ne lui ont pas paru différer essentiellement de ceux qu’il a publiés dans le journal de M. Jordan et les Annales de l’ENS.”

Rappelons que le traité de Bachelier de 1912 a joué un rôle important en Europe dans les années trente et aux Etats-Unis après la deuxième guerre mondiale...”
(d’après Bernard BRU, bulletin 60, octobre 1999, SMAI)

2.1 Le modèle

a) Soit $(W_t, t \geq 0)$ un mouvement brownien standard sur un espace de probabilité filtré $(\Omega, (\mathcal{F}_t), \mathbb{P})$ et $\xi_t = \log S_t$ où $d\xi_t = bdt + \sigma dW_t, \xi_0 = \log S_0$. Le processus S est observé aux instants déterministes $t_i = i\Delta, i = 1, \dots, N, N\Delta = T$. On cherche à estimer $\theta = (b, \sigma)$ dans $\mathbb{R} \times \mathbb{R}^+$. Si l’on considère ξ au lieu de S , on a un processus de Ito explicite et non plus une diffusion solution d’une EDS.

b) Le modèle statistique associé à cet ensemble d’observations est donné par
- l’espace des observations $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$,
- la famille de lois de probabilité du vecteur $(\xi_{i\Delta}, i = 1, \dots, N)$, indexée par le paramètre $\theta : (\mathbb{P}_\theta^N, \theta \in \mathbb{R} \times \mathbb{R}^+)$.
Ainsi, le modèle statistique est

$$[\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), (\mathbb{P}_\theta^N, \theta \in \mathbb{R} \times \mathbb{R}^+)].$$

Exercice : Ecrire la famille des lois. La famille est dominée par la mesure de Lebesgue. On obtient pour un N -échantillon de la trajectoire, puisque les accroissements sont indépendants :

$$\mathbb{P}_\theta^N(x) = \otimes_{i=1}^N \mathcal{N}(b\Delta, \sigma^2 \Delta)(x_{i+1} - x_i).$$

2.2 Vraisemblance

On observe que toutes les lois de cette famille sont dominées par la mesure de Lebesgue, on est donc dans les conditions d'application de la vraisemblance. On obtient

$$L(\theta) = \theta \mapsto \frac{1}{(\sigma\sqrt{2\pi\Delta})^N} \exp\left[-\frac{1}{2\sigma^2\Delta} \sum_{i=0}^{N-1} (\xi_{t_{i+1}} - \xi_{t_i} - b\Delta)^2\right].$$

Si l'on change de mesure dominante, par exemple $\nu = g.\mu$, la vraisemblance est multipliée par la densité de μ par rapport à ν , indépendante de θ , donc ne jouant aucun rôle au regard de la vraisemblance.

Dans le cas de ce modèle exponentiel, il est clair que l'on a tout intérêt à utiliser plutôt la log-vraisemblance que l'on note $l(\theta)$.

$$l(\theta) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2\Delta} \sum_{i=0}^{N-1} (\xi_{t_{i+1}} - \xi_{t_i} - b\Delta)^2.$$

Définition 2.1 On appelle estimateur de maximum de vraisemblance (EMV) tout élément de l'ensemble $\text{Arg max}L$.

Comme le log est croissant, $\text{Arg max}L = \text{Arg max}l$.

Exercice : rechercher les EMV de ce modèle

Les estimateurs qui annulent le gradient sont

$$\hat{b} = \frac{x_T^i - \xi_0}{T} ; \hat{a} = \frac{1}{T} \sum_i (x_{i+1} - x_i - \Delta\hat{b})^2.$$

Il faut vérifier que ce couple réalise un maximum de la fonction l ; d'une part l est concave en b et donc pour tout a \hat{b} réalise un maximum.

Ensuite, on a successivement :

$$\partial_a l = -\frac{N}{2a} + \frac{1}{2a^2\Delta} \sum_i (x_{i+1} - x_i - \Delta\hat{b})^2$$

et

$$\partial_a^2 l = \frac{N}{2a^2} - \frac{2}{2a^3\Delta} \sum_i (x_{i+1} - x_i - \Delta\hat{b})^2$$

et le tableau de variation en fonction de a montre que l'on a bien un maximum.

2.3 Variations quadratiques

Il n'y a pas d'unicité des estimateurs à proposer. Un estimateur classique de la volatilité est la variation quadratique

$$\tilde{\sigma}^2 = \frac{1}{T} \sum_{i=0}^{N-1} (\xi_{t_{i+1}} - \xi_{t_i})^2 = \hat{a} + (\hat{b})^2 \Delta$$

dont on sait ([20] ou [27]) qu'elle converge vers le crochet de la diffusion $\langle \xi \rangle_T$ et dont la loi (au coefficient σ^2/N près...) est celle d'un chi-deux (à n degrés de liberté) décentré de la quantité $\frac{b\sqrt{\Delta}}{\sigma}$ qui est petite en même temps que Δ . En effet, c'est la norme au carré d'un N -vecteur gaussien de variables aléatoires indépendantes de moyenne $\frac{b\Delta}{\sqrt{T}}$ et de variance $\frac{a\Delta}{T} = \frac{a}{N}$.

On peut chercher également des estimateurs en minimisant une erreur quadratique : les variables aléatoires X_{t_i} observées sont de loi connue et l'on peut calculer l'espérance de la distance quadratique

$$E[\sum_i (X_{t_i} - x_i)^2]$$

où x_i sont les valeurs observées. Calculons plutôt cette distance sur les écarts $X_{t_{i+1}} - X_{t_i}$ de loi gaussienne de moyenne $b\Delta$ et de variance $\sigma^2\Delta$:

$$E[\sum_i (b\Delta + \sigma\Delta_i W - (x_{t_{i+1}} - x_{t_i}))^2]$$

que l'on essaye de minimiser dans le couple (b, σ^2) .

On obtient comme précédemment \hat{b} pour estimateur de b mais la fonction à minimiser est croissante en a , le minimum est en 0 ce qui n'est ni intéressant ni raisonnable.

2.4 Propriétés des estimateurs obtenus

L'EMV obtenu est $(\frac{1}{\Delta}\bar{x}, \frac{1}{T} \sum_{i=0}^{N-1} (x_i - \bar{x})^2)$ où les observation sont $x_i = \xi_{t_{i+1}} - \xi_{t_i}$, et on recherche sa loi. Après simplification, il vient

$$\hat{b} = \frac{\xi_T}{T}, \text{ soit } b + \sigma \frac{W_T}{T},$$

$$\hat{a} = \frac{\sigma^2}{T} \sum_{i=0}^{N-1} (\Delta_i W - \frac{W_T}{N})^2.$$

Les X_i sont des variables aléatoires indépendantes, de loi de Gauss $(b\Delta, \sigma^2\Delta)$ et le théorème de Cochran montre alors que les deux estimateurs sont indépendants et que le couple est de loi $\mathcal{N}(b, \frac{\sigma^2}{T}) \otimes \chi_{N-1}^2$ à un coefficient multiplicatif près (σ^2/N) .

- Le biais de ce couple d'estimateurs est donc $(0, \sigma^2 \frac{-1}{N})$,

- l'erreur quadratique est $(\frac{\sigma^2}{T}, \sigma^4(\frac{2(N-1)^3 + N^2}{N^4}))$.

De fait, on peut montrer que $\hat{a} - \sigma^2$ converge presque sûrement vers 0 et que $\sqrt{N}(\hat{a} - \sigma^2)$ converge en loi vers une gaussienne centrée de variance $2\sigma^4$.

2.5 Asymptotique

(VGC I.1, pages 7.3,7.9)

Exercices

a) Lorsque T tend vers l'infini avec Δ fixe, on a donc obtenu des estimateurs asymptotiquement sans biais, convergents dans L^2 , et consistants grâce à la loi des grands nombres.

D'abord, $\hat{b} = b + \sigma \frac{W_T}{T}$, et les propriétés du mouvement brownien montrent que $\frac{W_T}{T}$ converge presque sûrement vers 0 : l'estimateur \hat{b} est consistant. De plus, l'erreur quadratique est en $1/T$: il y a donc également convergence dans L^2 et on a une convergence en \sqrt{T} .

Par ailleurs, on peut écrire

$$\hat{a} = \frac{a}{T} \sum_{i=1}^N (\Delta_i W - \frac{W_T}{N})^2,$$

qui peut encore se récrire

$$\hat{a} = \frac{a}{N} \sum_{i=1}^N (\frac{\Delta_i W}{\sqrt{\Delta}})^2 - a\Delta (\frac{W_T}{T})^2.$$

D'après l'étude de \hat{b} , le deuxième terme converge presque sûrement vers 0. De même, en utilisant la loi des grands nombres, on peut voir que le premier terme, moyenne de $i = 1$ à $i = N$ de variables aléatoires indépendantes deux à deux et intégrables, converge presque sûrement vers a , et donc \hat{a} est encore un estimateur consistant. Avec le théorème de la limite centrale, on peut de plus montrer une convergence en loi :

$$\sqrt{N}(\hat{a} - a) \rightarrow \mathcal{N}(0, 2a^2).$$

De plus, l'erreur quadratique est en $1/T$: il y a donc également convergence dans L^2 et on a une convergence en \sqrt{T} .

b) Lorsque Δ tend vers 0 avec T fixe, on a encore $N = T/\Delta$ tend vers l'infini, et les estimateurs sont asymptotiquement sans biais, $\hat{\sigma}$ converge dans L^2 , mais pas \hat{b} qui n'est pas non plus consistant : il a une loi fixe, et aucune convergence ne peut s'opérer....

En revanche, on peut sur \hat{a} mettre en évidence le même type de propriétés que pour le paragraphe a) : le deuxième terme avec Δ en facteur tend presque sûrement vers 0, et le premier converge encore presque sûrement vers a quand Δ tend vers 0, car alors N tend vers l'infini. De plus, on a encore la même convergence en loi.

$$\sqrt{T/\Delta}(\hat{a} - a) \rightarrow \mathcal{N}(0, 2a^2).$$

La conclusion est ici que, si l'on dispose d'une observation continue des prix sur $[0, T]$, alors a est complètement identifié.

c) Lorsque Δ tend vers 0 et T tend vers l'infini, N tend aussi vers l'infini, les estimateurs sont asymptotiquement sans biais, convergents dans L^2 , et consistants grâce à la loi des grands nombres :

\hat{b} converge presque sûrement vers b , l'erreur quadratique est en $1/T$: il y a donc également convergence dans L^2 et on a une convergence en \sqrt{T} et on a la convergence en loi :

$$\sqrt{T}(\hat{b} - b) \rightarrow \mathcal{N}(0, a).$$

Par ailleurs,

$$\hat{a} = \frac{a}{N} \sum_i \left(\frac{\Delta_i W}{\sqrt{\Delta}} \right)^2 - a \Delta \left(\frac{W_T}{T} \right)^2.$$

D'après l'étude de \hat{b} , le deuxième terme converge encore presque sûrement vers 0. Comme $T = N\Delta$ tend vers l'infini et Δ tend vers 0, nécessairement, N tend vers l'infini. Donc, en utilisant à nouveau la loi des grands nombres, on peut voir que le premier terme, moyenne de $i = 1$ à $i = N$ de variables aléatoires indépendantes deux à deux et intégrables, converge presque sûrement vers a , et donc \hat{a} est toujours un estimateur consistant. Avec le théorème de la limite centrale, on peut de plus montrer une convergence en loi :

$$\sqrt{N}(\hat{a} - a) \rightarrow \mathcal{N}(0, 2a^2).$$

De plus, l'erreur quadratique est en $1/T$: il y a donc également convergence dans L^2 et on a une convergence en \sqrt{T} .

La conclusion est ici qu'on estime simultanément de façon consistante a et b , et que l'on obtient des lois asymptotiques gaussiennes :

$$\sqrt{T}(\hat{b} - b) \text{ de loi } \mathcal{N}(0, 1) ; \sqrt{N}(\hat{a} - a) \rightarrow \mathcal{N}(0, 2a^2).$$

Notons qu'ils n'ont pas la même vitesse de convergence : \hat{a} est plus rapide.

d) Le cas de la petite variance se présente lorsque T est fixe mais σ connu tend vers 0. Alors, $\hat{b} = b + \sigma \frac{W_T}{T}$, $\hat{b} - b$ est de loi $\mathcal{N}(0, \frac{\sigma^2}{T})$ et $\frac{1}{\sigma}(\hat{b} - b)$ est de loi $\mathcal{N}(0, \frac{1}{T})$.

2.6

On peut enfin chercher des **intervalles de confiance**. On tire du fait que $\hat{b} - b$ suit une loi de Gauss $\mathcal{N}(0, \frac{\sigma^2}{T})$ l'encadrement

$$\mathbb{P}\{|\hat{b} - b| \leq C\} = \mathcal{N}(0, 1)\{|y| \leq C\sqrt{T/\sigma^2}\}$$

et au niveau 0.05, on obtient l'intervalle de confiance pour b

$$b \in [\hat{b} - 1.64 \frac{\sigma}{\sqrt{T}}, \hat{b} + 1.64 \frac{\sigma}{\sqrt{T}}].$$

Cet intervalle est intéressant aussi bien pour T grand que σ petit. On voit là que les résultats sont bons dans les deux cas : T grand, ou σ petit.

De même, $\frac{\hat{\sigma}^2 N}{\sigma^2}$ suivant une loi de chi-deux à $N - 1$ degrés de liberté, pour $N = 30$ par exemple, on obtient $\mathbb{P}\{16.8 \leq \frac{\hat{\sigma}^2 N}{\sigma^2} \leq 47\} = 0.95$, d'où l'intervalle de confiance

$$\sigma^2 \in [\hat{\sigma}^2 \frac{30}{47}, \hat{\sigma}^2 \frac{30}{16.8}].$$

2.7 Extensions

(VGC, I.1 pages(5-6))

De façon générale, on peut supposer qu'il existe un paramètre $\theta \in \mathbb{R}^k$, tel que

$$dX_t = b(t, \theta)dt + \sigma(t, \theta)dW_t,$$

les fonctions déterministes b et σ sont supposées connues et vérifient des hypothèses d'intégrabilité du type

$$\forall t \leq T, \forall \theta \in \Theta \int_0^t |b(s, \theta)|ds + \int_0^t \sigma^2(s, \theta)ds < +\infty.$$

On cherche à estimer θ . Dans ce cas, la vraisemblance s'écrit

$$L(\theta) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi \int_{t_i}^{t_{i+1}} \sigma^2(s, \theta)ds}} \exp\left[-\frac{(x_i - \int_{t_i}^{t_{i+1}} b(s, \theta)ds)^2}{2 \int_{t_i}^{t_{i+1}} \sigma^2(s, \theta)ds}\right].$$

La log-vraisemblance s'écrit :

$$l(\theta) = \sum_{i=0}^{N-1} -\frac{1}{2} \log \int_{t_i}^{t_{i+1}} \sigma^2(s, \theta) ds - \frac{(x_i - \int_{t_i}^{t_{i+1}} b(s, \theta) ds)^2}{2 \int_{t_i}^{t_{i+1}} \sigma^2(s, \theta) ds},$$

où x_i sont les valeurs observées pour les variables aléatoires $Y_i = X_{t_{i+1}} - X_{t_i}$ qui sont indépendantes et suivent des lois gaussiennes $\mathcal{N}(\int_{t_i}^{t_{i+1}} b(s, \theta) ds, \int_{t_i}^{t_{i+1}} \sigma^2(s, \theta) ds)$. Le calcul de l'EMV n'est pas forcément aisé selon la forme de b et σ ... Cependant, le calcul par une méthode numérique des EMV est possible dans la mesure où la vraisemblance est explicite.

Exercices : voir ce que cela donne avec

- . $b(t, \theta) = tm, \sigma(t, \theta) = \sigma, \theta = (m, \sigma)$.
- . $b(t, \theta) = mf(t), \sigma(t, \theta) = \sigma, \theta = (m, \sigma)$.
- . $b(t, \theta) = m, \sigma(t, \theta) = \sigma\sqrt{1+t^2}, \theta = (m, \sigma)$.

Premier exemple :

$$l(\theta) = \sum_{i=0}^{N-1} \left[-\frac{1}{2} \log(\sigma^2 h) - \frac{1}{2} \frac{(Y_i - m \frac{1}{2} h^2 (2i+1))^2}{h \sigma^2} \right].$$

Cette application est concave en m et on obtient l'EMV

$$\hat{m} = \frac{\sum_{i=0}^{N-1} X_i (i + \frac{1}{2})}{h^2 \sum_{i=0}^{N-1} (i + \frac{1}{2})^2}.$$

C'est un estimateur sans biais de variance $\frac{\sigma^2}{h}$.

On peut montrer que le suivant est aussi EMV :

$$\hat{a} = \frac{1}{T} \sum_{i=0}^{N-1} (Y_i - (i + \frac{1}{2}) \hat{m} h^2)^2.$$

2.8 Simulations

Un mot en conclusion sur les simulations : lorsque l'on ne dispose pas de données réelles, ou pour valider la méthode d'estimation que l'on a choisie, il est intéressant de simuler des données. Ceci est particulièrement vrai lorsqu'on sait faire une simulation exacte de la trajectoire observée. C'est le cas des modèles de processus à accroissements indépendants gaussiens. Soit à simuler $(\xi_t, i = 1, N)$ lorsque la trajectoire ξ est définie par :

$$d\xi_t = b(t, \theta) dt + \sigma(t, \theta) dW_t, \xi_0 = 0.$$

Si $(\varepsilon_1, \dots, \varepsilon_N)$ est un N -échantillon de la loi $\mathcal{N}(0, 1)$, il suffit d'écrire

$$\xi_{t_{i+1}} = \xi_{t_i} + \int_{t_i}^{t_{i+1}} b(s, \theta) ds + \sqrt{\int_{t_i}^{t_{i+1}} \sigma^2(s, \theta) ds} \varepsilon_i,$$

pour $i = 1, \dots, N$. Puis l'on teste les estimateurs selon l'algorithme suivant :

- on fixe θ ,
- on simule N -échantillon de la loi $\mathcal{N}(0, 1)$: $(\varepsilon_1, \dots, \varepsilon_N)$,
- on calcule $(\xi_{t_1}, \dots, \xi_{t_N})$ en fonction de la valeur de θ choisie et des valeurs obtenues pour le N -échantillon,
- on calcule l'estimation $\theta(\xi_{t_1}, \dots, \xi_{t_N})$
- que l'on compare au θ choisi en première étape.

Ceci permet de tester la qualité de l'estimation choisie.

3 Modèle de Vasicek

Ce modèle correspond à celui appelé de “Ornstein-Uhlenbeck” par les probabilistes :

$$(5) \quad dX_t = \theta X_t dt + \sigma dW_t, X_0 = \eta,$$

où η est une variable aléatoire indépendante du mouvement brownien W , \mathcal{F}_0 -mesurable. Il sert par exemple à modéliser une volatilité aléatoire (cf le cours de Gaudron : équation de Langevin).

3.1 Existence d’une solution

L’équation (5) est une équation différentielle stochastique de paramètres à estimer $(\theta, \sigma) \in \mathbb{R} \times \mathbb{R}^+$.

Proposition 3.1 *L’équation (5) admet l’unique solution*

$$X_t = e^{\theta t} \eta + \sigma e^{\theta t} \int_0^t e^{-\theta s} dW_s.$$

De plus, pour tout $t \geq 0$, pour tout $\Delta \geq 0$,

$$X_{t+\Delta} - e^{\theta \Delta} X_t = \sigma e^{\theta(t+\Delta)} \int_t^{t+\Delta} e^{-\theta s} dW_s$$

et la variable aléatoire $Z_{t,\Delta} = X_{t+\Delta} - e^{\theta \Delta} X_t$ est de loi gaussienne $\mathcal{N}(0, \sigma^2 \frac{e^{2\theta \Delta} - 1}{2\theta})$ indépendante de X_t .

Preuve : (i) l’unicité est conséquence de la théorie des équations différentielles stochastiques car les coefficients sont lipschitziens : $b(x) = \theta x, \sigma(x) = \sigma$. Mais plus simplement, si X^1 et X^2 sont solutions,

$$d(X_t^1 - X_t^2) = \theta(X_t^1 - X_t^2)dt, (X_0^1 - X_0^2) = 0,$$

qui admet l’unique solution identiquement nulle sur \mathbb{R}^+ .

(ii) Le fait que la solution est de cette forme, une fois l’unicité établie, est conséquence de la dérivation de Itô de la forme proposée ; il s’agit d’une simple vérification.(exercice)

(iii) On écrit les deux termes de la différence cherchée :

$$\begin{aligned} X_{t+\Delta} &= e^{\theta \Delta} e^{\theta t} \eta + \sigma e^{\theta \Delta} e^{\theta t} \left[\int_0^t e^{-\theta s} dW_s + \int_t^{t+\Delta} e^{-\theta s} dW_s \right], \\ e^{\theta \Delta} X_t &= e^{\theta \Delta} (e^{\theta t} \eta + \sigma e^{\theta t} \int_0^t e^{-\theta s} dW_s), \end{aligned}$$

et on simplifie. Puisque θ et Δ sont réels, on a clairement que la variable aléatoire $Z_{t,\Delta}$ suit une loi gaussienne, centrée, de variance $\sigma^2 e^{2\theta(t+\Delta)} \int_t^{t+\Delta} e^{-2\theta s} ds$ qui se calcule aisément.

□

Remarque 3.2 Si η est gaussienne, indépendante de W , X_t est aussi gaussienne, de loi $\mathcal{N}(e^{\theta t}E[\eta], e^{2\theta t}\sigma_\eta^2 + \sigma^2\frac{e^{2\theta t}-1}{2\theta})$. De plus, X est alors un processus gaussien, de fonction de covariance

$$\text{cov}(X_t, X_{s+t}) = e^{\theta s} [e^{2\theta t}\sigma_\eta^2 + \sigma^2\frac{e^{2\theta t}-1}{2\theta}].$$

Preuve : (i) pour tout k , tout k -uple croissant (t_1, \dots, t_k) , le vecteur $(X_{t_1}, \dots, X_{t_k})$ est de loi image par F linéaire de la loi $\mathbb{P}^\eta \otimes_1^k \mathcal{N}(0, \sigma^2\frac{e^{2\theta(t_{i+1}-t_i)}-1}{2\theta})$. Comme F est linéaire (de fait sa matrice est triangulaire avec des 1 sur la diagonale, et le jacobien vaut 1) la loi image est gaussienne, et le processus est gaussien.

$$F : (y_i) \mapsto (y_i - e^{-\theta\Delta_i t}).$$

(ii) $X_{t+s} = e^{\theta s}X_t + X_{t+s} - e^{\theta s}X_t$ comme somme de deux variables aléatoires indépendantes. La covariance cherchée est donc $e^{\theta s}\text{Var}(X_t)$

- Si $\theta < 0$ et $\sigma_\eta^2 = -\frac{\sigma^2}{2\theta}$, cette covariance est $-e^{\theta s}\frac{\sigma^2}{2\theta}$.

- Si η est centrée, $\theta < 0$ et $\sigma_\eta^2 = -\frac{\sigma^2}{2\theta}$, on obtient un processus gaussien stationnaire.

3.2 Discrétisation

On obtient par discrétisation un AR(1). Soit $t_i = ih, i \in \mathbb{N}, \eta$ variable aléatoire indépendante du mouvement brownien :

$$X_{(i+1)h} = e^{\theta h}X_{ih} + \sigma e^{\theta(i+1)h} \int_{ih}^{(i+1)h} e^{-\theta s} dW_s.$$

Donc, si l'on pose $Y_i = X_{ih}$, Y_{i+1} est $e^{\theta h}Y_i$ plus une gaussienne indépendante, centrée, de variance $\sigma^2\frac{e^{2\theta h}-1}{2h}$ et l'on peut considérer que (Y_i) est une suite solution de l'équation ARMA :

$$Y_{i+1} = e^{\theta h}Y_i + \sigma\sqrt{\frac{e^{2\theta h}-1}{2\theta}}\varepsilon_{i+1}$$

où la suite ε est un bruit blanc.

Corollaire 3.3 La suite $(Y_i, i \in \mathbb{N})$ est un AR(1).

On est donc ramené pour estimer θ et σ à l'estimation des paramètres de l'AR(1) :

$$a_1 = e^{\theta h}, b_0 = \sigma\sqrt{\frac{e^{2\theta h}-1}{2h}}.$$

On peut utiliser les algorithmes de Box et Jenkins ou n'importe quel logiciel correct de séries chronologiques. On a même dans ce cas un test du modèle par exemple sur les résidus après avoir effectué les estimations.

bien sûr, on a les expressions explicites des estimateurs, mais c'est qd même agréable de récupérer ce que autrui a déjà programmé.... θ s'estime par $\frac{1}{h} \log \hat{a}_1$ et σ s'estime par $\hat{b}_0 \sqrt{\frac{2h}{\hat{a}_1^2 - 1}}$.

3.3 Fonction de vraisemblance du modèle discrétisé

On suppose ici que $\eta = x$ déterministe dans \mathbb{R} et connu. Les observations se font aux temps $ih, i = 0, \dots, N; T = hN$). Le modèle statistique est alors $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mathbb{P}^{\theta, \sigma})$ où la densité de $\mathbb{P}^{\theta, \sigma}$ par rapport à la mesure de Lebesgue est donnée par ($a = \sigma^2$) :

$$L(\theta, a) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi \frac{e^{2h\theta}-1}{2\theta} a}} \exp - \frac{(x_i - e^{\theta h} x_{i-1})^2}{2 \frac{e^{2h\theta}-1}{2\theta} a}$$

(obtenue à l'aide de la remarque 3.2) et la log-vraisemblance

$$-\frac{1}{2} [\log a^N + \log v(\theta)^N + \sum_{i=1}^N \frac{(x_i - e^{\theta h} x_{i-1})^2}{\frac{e^{2h\theta}-1}{2\theta} a}]$$

où $v(\theta) = \frac{e^{2h\theta}-1}{2\theta}$.

Exercice : rechercher les estimateur de maximum de vraisemblance de a et $v(\theta)$

On fait le changement de paramètres : $\theta_1 = e^{h\theta}$ (c'est le paramètre a_1 de l'AR(1)) ; $\theta_2 = av(\theta)$ (c'est le paramètre b_0^2 de l'AR(1)). Ce changement est une bijection monotone et il suffit de maximiser l'application

$$L : (\theta_1, \theta_2) \mapsto -\frac{1}{2} N \log \theta_2 - \frac{1}{2\theta_2} \sum_i (x_i - \theta_1 x_{i-1})^2.$$

On obtient :

$$\hat{\theta}_1 = \frac{\sum_i x_i x_{i-1}}{\sum_i x_{i-1}^2} ; \hat{\theta}_2 = \frac{1}{N} \sum_i (x_i - \hat{\theta}_1 x_{i-1})^2.$$

Ceci se justifie par l'étude de L . Le gradient de L est

$$\nabla_{\theta_1} l = \frac{1}{\theta_2} \sum_i x_{i-1} (x_i - \theta_1 x_{i-1}); \nabla_{\theta_2} l = -\frac{N}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_i (x_i - \theta_1 x_{i-1})^2.$$

L est concave en sa première variable. La même étude que celle menée en 2.2 conduit au résultat.

Il faut remarquer que $\hat{\theta}^1$ est censé estimer une exponentielle, mais n'est pas forcément positif : si l'on exprime l'estimateur, il vient au numérateur (le dénominateur est positif) :

$$\sum_i (e^{h\theta} X_{(i-1)h} + Z_{(i-1)h,h}) X_{(i-1)h}$$

dont la loi est connue : la variance de Z est négligeable devant celle de X , aussi la probabilité que certains des termes de la somme soient négatifs est faible. Donc, la probabilité que la somme est négative est faible.

3.4 Propriétés

3.4.1 AR(1)

D'après la discrétisation, on tire évidemment les estimations des coefficients de l'équation ARMA et toutes les propriétés classiques. Voir le cours de M. Garel.

3.4.2 estimateur de maximum de vraisemblance

Les lois exactes des estimateur de maximum de vraisemblance sont difficiles à calculer.... On cherche d'autres estimateurs plus faciles à étudier, même s'ils n'ont plus la propriété de maximiser la vraisemblance.

3.5 Estimateurs "naturels"

Par exemple, si h est petit, $v(\theta)$ est proche de h , la log vraisemblance est proche de

$$-\frac{N}{2} \log a - \frac{1}{2ah} \sum_i (x_i - x_{i-1} - \theta h x_{i-1})^2,$$

et on peut penser aux estimations suivantes de $a = \sigma^2$ et de θ (en négligeant $\theta h x_{i-1}$) :

$$\begin{aligned}\tilde{a} &= \frac{1}{Nh} \sum_i (x_i - x_{i-1})^2, \\ \tilde{\theta} &= \frac{\sum_{i=1}^N x_{i-1} (x_i - x_{i-1})}{h \sum_{i=1}^N x_{i-1}^2}.\end{aligned}$$

(Pour justifier $\tilde{\theta}$, écrire $\hat{\theta} = \frac{1}{h} \log(\hat{\theta}^1) = \frac{1}{h} \log(1 + h\tilde{\theta}) \sim \tilde{\theta}$.)

Les formules donnant $\tilde{\theta}$ et \tilde{a} définissent d'autres estimateurs que l'on peut étudier également. Ainsi, lorsque T est fixe et N tend vers l'infini, le pas h tend vers 0, et presque sûrement, au sens des intégrales de Stieljes, $\tilde{\theta}$ converge vers

$$\frac{\int_0^T X_s dX_s}{\int_0^T X_s^2 ds}.$$

Quant à \tilde{a} , c'est la somme de trois termes :

$$T\tilde{a} = \theta^2 \sum_{i=1}^N \left(\int_{I_i} X_s ds \right)^2 + 2\theta\sigma \sum_{i=1}^N \int_{I_i} X_s ds \Delta_i W + \sigma^2 \sum_{i=1}^N (\Delta_i W)^2.$$

Presque sûrement, le premier terme est un $O(h)$, le second un $O(\sqrt{h})$ et le dernier converge presque sûrement vers $\sigma^2 T$. On a donc un estimateur consistant de a . On reprendra ceci plus en détail au paragraphe 3.6.3

3.6 Asymptotique

3.6.1 Comportement de X_t lorsque t tend vers l'infini

VGC pages 11 13 et DDC-MD pages 80 et sq....

La proposition suivante décrit le comportement asymptotique de la variable aléatoire X_t lorsque t tend vers l'infini.

Proposition 3.4 :

- (i) Si $\theta < 0$, X_t converge en loi lorsque t tend vers l'infini vers la loi $\mathcal{N}(0, \frac{a}{-2\theta})$.
(ii) Si $\theta > 0$, $X_t e^{-\theta t}$ converge dans $L^2(\mathbb{P})$ et presque sûrement vers $Z_\theta = \eta + \sigma \int_0^\infty e^{-\theta s} dW_s$.
Si η est de loi $\mathcal{N}(m, d)$, indépendante de W , la variable aléatoire Z_θ a la loi $\mathcal{N}(m, d + \frac{a}{2\theta})$. ($a = \sigma^2$).

Preuve : Dans les deux cas, on a

$$X_t = e^{\theta t} \eta + \sigma e^{\theta t} \int_0^t e^{-\theta s} dW_s.$$

(i) Si $\theta < 0$, $e^{\theta t} \eta$ converge presque sûrement vers 0, donc en probabilité, donc en loi. Le deuxième terme est une variable aléatoire gaussienne ; sa convergence en loi est assurée dès que son espérance et sa variance converge. Or, l'espérance est nulle et la variance est $a \frac{e^{2\theta t} - 1}{2\theta}$. De plus, $\lim_{t \rightarrow \infty} a \frac{e^{2\theta t} - 1}{2\theta} = \frac{a}{-2\theta}$.

(ii) Si $\theta > 0$, $X_t e^{-\theta t} = \eta + \sigma \int_0^t e^{-\theta s} dW_s$. Il suffit d'examiner la limite de $\|\int_t^\infty e^{-\theta s} dW_s\|_2^2$. Or cette norme au carré vaut $\int_t^\infty e^{-2\theta s} ds = e^{-2\theta t}$ dont la limite est nulle à l'infini, d'où le résultat.

Remarquons de plus que $t \mapsto \int_0^t e^{-\theta s} dW_s$ est une martingale ; si elle converge dans L^2 , elle converge dans L^1 , elle est donc régulière et converge presque sûrement vers $\int_0^\infty e^{-\theta s} dW_s$. \square

Nous avons décrit le comportement asymptotique de X_t pour t infini ; ceci sera utilisé pour le comportement asymptotique des estimateurs de la section 3.5 obtenus sur la base d'une observation continue de X sur tout l'intervalle de temps.

3.6.2 σ tend vers 0 et T fixe

Que se passe-t-il lorsque la variance est connue et tend vers 0 ?

Proposition 3.5 Pour tout $T > 0$, tout $\theta \in \mathbb{R}$, on obtient

$$\sup_{0 \leq t \leq T} |X_t - e^{\theta t} \eta| \rightarrow 0$$

dans tout L^p et presque sûrement lorsque σ tend vers 0.

Preuve : $X_t - e^{\theta t} \eta = \sigma e^{\theta t} \int_0^t e^{-\theta s} dW_s$. Or, par les propriétés du mouvement brownien, la variable aléatoire

$$\sup_{0 \leq t \leq T} e^{\theta t} \left| \int_0^t e^{-\theta s} dW_s \right|$$

est bien définie presque sûrement. Par conséquent, la convergence se déduit grâce au facteur σ . Plus précisément, par les inégalités de Burkholder, on contrôle tous les moments de la martingale $\int_0^t e^{-\theta s} dW_s$: il existe une constante $C_p > 0$ telle que :

$$\left\| \sup_{0 \leq t \leq T} \left| \int_0^t e^{-\theta s} dW_s \right| \right\|_p \leq C_p \left\| \int_0^T e^{-\theta s} dW_s \right\|_p, \forall p \geq 1.$$

3.6.3 T fixe et $h = \frac{T}{N}$ tend vers 0

On revient ici sur les estimateur de maximum de vraisemblance du paragraphe 3.5, estimateurs “naturels”.

$$\begin{aligned} \tilde{a} &= \frac{1}{Nh} \sum_i (x_i - x_{i-1})^2, \\ \tilde{\theta} &= \frac{\sum_{i=1}^N x_{i-1} (x_i - x_{i-1})}{h \sum_{i=1}^N x_{i-1}^2}. \end{aligned}$$

(a) D’après le schéma de discrétisation $X_{ih} - X_{(i-1)h} = \theta \int_{(i-1)h}^{ih} X_s ds + \sigma(W_{ih} - W_{(i-1)h})$.
Donc au carré :

$$(X_{ih} - X_{(i-1)h})^2 = \theta^2 \left(\int_{(i-1)h}^{ih} X_s ds \right)^2 + 2\sigma\theta(W_{ih} - W_{(i-1)h}) \int_{(i-1)h}^{ih} X_s ds + \sigma^2(W_{ih} - W_{(i-1)h})^2.$$

Le premier terme est majoré par $\theta^2 h^2 (X_T^*)^2$, de moyenne de $i = 1$ à N majorée par $\frac{1}{Nh} N \theta^2 h^2 (X_T^*)^2$, qui est un $O(h)$ (on note $X_T^* = \sup_{0 \leq s \leq T} |X_s|$ qui est presque sûrement fini parce que X est continue sur $[0, T]$).

Le carré de la somme des doubles produits par Cauchy-Schwartz des intégrales discrètes est majoré par le produit de la somme des carrés, soit

$$\frac{1}{Nh} \left| \sum_i \sigma\theta(W_{ih} - W_{(i-1)h}) \int_{(i-1)h}^{ih} X_s ds \right| \leq \frac{1}{Nh} \sqrt{N\theta^2 h} X_T^* \sqrt{\sum_i \sigma^2 (W_{ih} - W_{(i-1)h})^2} = O(\sqrt{h}).$$

La somme de carré d’accroissements de browniens converge vers T presque sûrement par la loi des grands nombres.

Globalement, l’estimateur \tilde{a} converge vers σ^2 presque sûrement lorsque h tend vers 0.

On peut montrer de plus, avec des outils supplémentaires :

$$\sqrt{N}(\tilde{a} - \sigma^2) \text{ converge en loi vers } \mathcal{N}(0, 2\sigma^2)$$

(b) On montre que $\tilde{\theta}_N$ converge presque sûrement vers $\hat{\theta}_T = \frac{\int_0^T X_s dX_s}{\int_0^T X_s^2 ds}$.

3.6.4 T tend vers l'infini et $h = \frac{T}{N}$ tend vers 0

Lorsque T tend vers l'infini, on peut étudier le comportement de l'estimateur $\tilde{\theta}$. On suppose pour simplifier que $X_0 = 0$ et $\sigma = 1$.

L'estimateur $\hat{\theta}_T = \frac{\int_0^T X_s dX_s}{\int_0^T X_s^2 ds}$ a quelques bonnes propriétés, mais difficiles à montrer. On étudie la différence :

$$\hat{\theta}_T - \theta = \frac{\int_0^T X_s (dX_s - \theta X_s ds)}{\int_0^T X_s^2 ds} = \frac{\int_0^T X_s dW_s}{\int_0^T X_s^2 ds}.$$

(i) Si $\theta < 0$, on peut montrer que $\sqrt{T}(\hat{\theta}_T - \theta)$ converge en loi vers une gaussienne $\mathcal{N}(0, -2\theta)$. (référence : un article de 1985, Mishra et Prakasa-Rao ; on utilise TCL et martingales)

(ii) Lorsque $\theta > 0$, [13] montrent que

$$\sqrt{\frac{e^{2\theta T} - 1}{2\theta}}(\hat{\theta}_T - \theta) \text{ converge en loi vers } U/Z$$

où (U, Z) suit une loi $\mathcal{N}(0, 1) \otimes \mathcal{N}(0, 1/2\theta)$ (VGC pages 17 bis et ter).

Esquisse de preuve (voir aussi [7] dans les pages 80) : on sait que dans ce cas $X_t e^{-\theta t}$ converge dans L^2 pour t infini vers $\int_0^\infty e^{-\theta s} dW_s$ et suit une loi normale centrée de variance $\frac{e^{2\theta t} - 1}{2\theta}$, notée $m_t(\theta)$.

. On montre $(m_t(\theta))^{-1} \int_0^t X_s^2 ds$ converge dans L^1 pour t infini vers $Z_\theta = (\int_0^\infty e^{-\theta s} dW_s)^2$. En effet,

$$\int_0^t X_s^2 ds - Z_\theta m_t(\theta) = \int_0^t (X_s^2 e^{-2\theta s} - Z_\theta) e^{2\theta s} ds$$

et pour tout ε , il existe $T, \forall t \geq T, \|e^{-\theta t} X_t - \int_0^\infty e^{-\theta s} dW_s\|_2 \leq \varepsilon$ soit :

$$\|m_t(\theta)^{-1} \int_0^t X_s^2 ds - Z_\theta\|_1 \leq m_t(\theta)^{-1} \left[\left\| \int_0^T X_s^2 ds \right\|_1 + \|Z_\theta\|_1 \int_0^T e^{2\theta s} ds + \varepsilon \int_T^\infty \|e^{-\theta s} X_s + Z_\theta^{\frac{1}{2}}\|_2 e^{2\theta s} ds \right].$$

.. On peut montrer (mais c'est long et difficile : 3-4 pages...) que le couple

$$(m_t(\theta)^{-\frac{1}{2}} \int_0^t X_s dW_s, m_t(\theta)^{-1} \int_0^t X_s^2 ds)$$

converge en loi vers $(U\sqrt{Z_\theta}, Z_\theta)$, où U et Z_θ sont indépendantes et U gaussienne, centrée, réduite.

Ce dernier point implique la convergence cherchée

3.7 Extensions

VGC pages 18-20

Supposons que l'on observe le processus défini par

$$(6) \quad dX_t = (a(t, \theta)X_t + b(t, \theta))dt + \sigma(t, \theta)dW_t, X_0 = \eta,$$

où η est une variable aléatoire indépendante du mouvement brownien W . Les fonctions a, σ et b sont déterministes et connues, sur $\mathbb{R}^+ \times \Theta$ et vérifient

$$\int_0^t |a(s, \theta)|ds + \int_0^t |b(s, \theta)|ds + \int_0^t \sigma^2(s, \theta)ds < \infty$$

pour tout $(t, \theta) \in \mathbb{R}^+ \times \Theta$, ce qui assure l'existence d'une unique solution forte de l'équation différentielle stochastique (6).

On peut faire la même étude que précédemment ; toutefois, les calculs et les expressions obtenues sont plus complexes.

Exercice : trouver la forme de la solution (méthodes de variation des constantes). Posons $a_t = a(t, \theta), b_t = b(t, \theta), \sigma_t = \sigma(t, \theta)$ où θ est le vrai paramètre. L'EDS (6) se résout en

$$X_t = \exp\left(\int_0^t a_s ds\right) \left[\eta + \int_0^t b_s e^{-\int_0^s a_u du} ds + \int_0^t \sigma_s e^{-\int_0^s a_u du} dW_s \right].$$

C'est une simple vérification par la formule de Itô (l'unicité obtenue grâce aux hypothèses sur (a, b, σ)).

Soit

$$\begin{aligned} \alpha_i &= \exp\left(\int_{t_{i-1}}^{t_i} a_s ds\right), \\ m_i &= \exp\left(\int_0^{t_i} a_s ds\right) \times \left[\int_{t_{i-1}}^{t_i} b_s e^{-\int_0^s a_u du} ds \right], \\ \sigma_i^2 &= \exp\left(2 \int_0^{t_i} a_s ds\right) \times \left[\int_{t_{i-1}}^{t_i} \sigma_s^2 e^{-2 \int_0^s a_u du} ds \right]. \end{aligned}$$

Exercice : montrer que $X_{t_i} = \alpha_i X_{t_{i-1}} + m_i + \sigma_i \varepsilon_i$ où $(\varepsilon_i, i = 1, \dots, N)$ est un N -échantillon de la loi gaussienne centrée réduite.

Ces formules permettent de calculer la densité du vecteur $(X_{t_i}, i = 1, \dots, N)$ et également de *simuler* cette trajectoire.

Exercice : Appliquer ce qui précède au modèle

$$(7) \quad dX_t = \theta f_t X_t dt + \sigma dW_t, X_0 = 0.$$

Calculer les estimateur de maximum de vraisemblance du paramètre (θ, σ^2) fondés sur l'observation du vecteur $(X_{t_i}, i = 1, \dots, N)$ lorsque $t_i = ih, i = 1, \dots, N$.

On pose F telle que $F' = f$; on a

$$a(t, \theta) = \theta f_t ; b = 0 ; \sigma(t, \theta) = \sigma$$

$$\alpha_i = e^{\theta \int_{I_i} f_s ds} ; m_i = 0; \sigma_i^2 = e^{2\theta F(t_i)} \sigma^2 \int_{I_i} e^{-2\theta F(s)} ds.$$

Et la log-vraisemblance s'écrit :

$$\prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2} \frac{(x_i - \alpha_i x_{i-1})^2}{\sigma_i^2}\right].$$

Là encore, pour mener les calculs à bout, on peut supposer la discrétisation assez fine pour remplacer les exponentielles de petit exposant par 1 plus l'exposant petit.

Pour conclure, on peut ajouter que ce qui vient d'être fait pour ces deux classes d'exemples ne peut être reproduit pour d'autres modèles. En effet, le trait particulier commun aux deux exemples est que la loi jointe du vecteur $(X_{t_i}, i = 1, \dots, N)$ est connue de façon explicite, ce qui n'est pas le cas pour des modèles généraux de diffusion. D'où la nécessité d'introduire un nouvel outil : le contraste.

4 Schéma d'Euler et fonction de contraste

4.1 Introduction, définitions

Dans des cas plus généraux que les deux modèles précédemment traités, la vraisemblance exacte n'est pas explicite. Une alternative est de discrétiser le modèle par un schéma d'Euler. On substitue à la vraisemblance du modèle celle du schéma discrétisé (qui est un processus de Markov) appelée alors **contraste**.

4.1.1 Le modèle

$$(8) \quad dX_t = b(\theta, t, X_t)dt + \sigma(\theta, t, X_t)dW_t, X_0 = \eta, \theta \in \Theta \subset \mathbb{R}^p.$$

Le problème est d'estimer θ étant données N observations du processus : $(X_{ih}, i = 1, \dots, N)$, les observations notées comme d'habitude (x_i) .

4.1.2 Le schéma d'Euler

Prenons d'abord un exemple simple où $\theta = (c, \sigma)$, avec $b(\theta, t, X_t) = b(c, X_t)$; $\sigma(\theta, t, X_t) = \sigma$. Le schéma d'Euler est

$$(9) \quad \tilde{X}_{ih} = \tilde{X}_{(i-1)h} + b(c, \tilde{X}_{(i-1)h})h + \sigma(W_{ih} - W_{(i-1)h}), X_0 = \eta, \theta \in \Theta \subset \mathbb{R}^2.$$

Remarquer que \tilde{X}_{ih} n'est pas exactement X_{ih} !!

Une étude de la différence $X - \tilde{X}$ où $\tilde{X}_s = X_{h[s/N]}$ et le lemme de Gronwall montre la convergence attendue.

La chaîne de Markov Y , ($Y_i = \tilde{X}_{ih}$) solution de cette équation est une chaîne de Markov homogène de noyau de transition paramétré par θ :

$$p_h(x, y; \theta) = \mathcal{N}(x + hb(c, x), \sigma^2 h).$$

La vraisemblance de ce processus de Markov se calcule car la loi du vecteur (Y_1, \dots, Y_N) est la loi de $Y_1 \times$ la loi de $Y_2/Y_1 \times$ etc \times la loi de Y_N sachant la loi de Y_{N-1} à savoir :

$$L(\theta; y_1, \dots, y_N) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2 h}} \exp\left[-\frac{(y_i - y_{i-1} - hb(c, y_{i-1}))^2}{2\sigma^2 h}\right]$$

et l'opposé de la log-vraisemblance définit alors le contraste :

$$C(\theta; y_1, \dots, y_N) = \frac{1}{2}N \log \sigma^2 + \frac{1}{2\sigma^2 h} \sum_{i=1}^N (y_i - y_{i-1} - hb(c, y_{i-1}))^2.$$

On pose $a = \sigma^2$. L'étude des estimateurs de minimum de contraste (EMC) donne

$$(10) \quad \tilde{c} = \arg \inf \sum_{i=1}^N (y_i - y_{i-1} - hb(c, y_{i-1}))^2,$$

$$(11) \quad \tilde{a} = \frac{1}{N} \sum_{i=1}^N (y_i - y_{i-1} - hb(\tilde{c}, y_{i-1}))^2,$$

ce qui réalise bien le maximum de C par une étude analogue à celle déjà vue.

$$\begin{aligned}\partial_a l &== -\frac{N}{2a} + \frac{\sum_{i=1}^N (y_i - y_{i-1} - hb(\tilde{c}, y_{i-1}))^2}{2a^2 h}, \\ \partial_{a^2}^2 l &== \frac{N}{2a^2} - \frac{\sum_{i=1}^N (y_i - y_{i-1} - hb(\tilde{c}, y_{i-1}))^2}{a^3 h}\end{aligned}$$

et le tableau de variations montre que \tilde{a} est bien un EMV.

Pour l'optimisation de $c \mapsto C(a, c), (y_i)$, cela dépend de la fonction b ; on peut opérer avec du calcul numérique. Il reste deux problèmes théoriques : l'existence d'un optimum et la vitesse de convergence.

Les résultats ci-dessous ne sont pas vraiment rédigés quelque part, et l'étude de ces estimateurs de minimum de contraste est assez délicate.

Valentine Genon-Catalot :” il faut distinguer deux cas
- d'abord $\Delta_n = T/n$ avec T fixe et n tendant vers l'infini ; dans ce cas, on estime les paramètres du coefficient de diffusion. Les estimateurs sont consistants, leur loi est asymptotiquement un mélange de gaussiennes (ce sont des théorèmes limites très difficiles) ;
- cas où Δ_n tend vers 0 et $T_n = \Delta_n n$ tend vers l'infini : là, il faut avoir des hypothèses d'ergodicité sur le modèle qui dépendent de l'étude du processus X et ensuite, il faut des théorèmes spécifiques, pas simples, qui sont l'objet de thèses et d'articles récents.”

Proposition 4.1 *Si $a = \sigma^2$ est connu, on a la convergence en loi lorsque N tend vers l'infini :*

$$\sqrt{N}(\tilde{c}_N - c) \rightarrow \mathcal{N}(0, V_h(c)).$$

Par ailleurs, pour la variance, si l'on suppose h petit et b bornée, on peut prendre plutôt l'estimateur dit “naturel” :

$$\tilde{a}_N^1 = \frac{1}{T} \sum_{i=1}^N (Y_i - Y_{i-1})^2$$

et on peut prouver

Proposition 4.2 *On a la convergence en loi lorsque N tend vers l'infini, T fixe, $T = Nh$:*

$$\sqrt{N}(\tilde{a}_N^1 - a) \rightarrow \mathcal{N}(0, 2a^2).$$

Preuve : $\tilde{a}_N^1 = \frac{1}{T} \sum_{i=1}^N (Y_i - Y_{i-1})^2$ est d'espérance $E[\tilde{a}_N^1] = \frac{1}{T} \sum_{i=1}^N h^2 E[b^2(\tilde{c}, Y_{i-1})] + a$ qui converge vers a par exemple si b est bornée et l'estimateur est asymptotiquement sans biais.

Par ailleurs, cette variable aléatoire vérifie :

$$\sqrt{N}(\tilde{a}_N^1 - a) = \sqrt{N} \left[\frac{1}{Nh} \sum_{i=1}^N ((Y_i - Y_{i-1})^2 - ha) \right]$$

dont chaque terme peut se récrire

$$(Y_i - Y_{i-1})^2 - ha = b^2(c, Y_{i-1})h^2 + a(\Delta_i W^2 - h) + 2hb(c, Y_{i-1})\sigma \Delta_i W$$

d'où il vient

$$\sqrt{N}(\tilde{a}_N^1 - a) = \frac{\sqrt{N}}{hN} \left[h^2 \sum_{i=1}^N b^2(c, Y_{i-1}) + a \left(\sum_{i=1}^N \Delta_i W^2 - T \right) + 2h\sigma \sum_{i=1}^N b(c, Y_{i-1}) \Delta_i W \right]$$

Si b est borné, le premier terme est un $O(\sqrt{h})$, le deuxième par le théorème de la limite centrale converge en loi vers une loi normale :

$$\frac{\sum_i (\Delta_i W)^2 - T}{h\sqrt{2N}} \rightarrow \mathcal{N}(0, 1),$$

et le troisième est analogue à celui d'une précédente étude : il ne peut être brutalement majoré par la racine du produit des somme des carrés qui est un $O(1)$. On opère de façon plus subtile. Le coefficient de $2h\sigma$ est une somme qui, par définition de l'intégrale stochastique, converge en probabilité vers une intégrale :

$$\sum_{i=1}^N b(c, x_{i-1}) \Delta_i W \rightarrow \int_0^T b(c, X_s) dW_s.$$

Ce terme multiplié alors par $\frac{\sqrt{N}}{hN} 2h\sigma = 2\sigma \frac{1}{\sqrt{N}}$ converge donc en probabilité, donc en loi, vers 0.

4.2 Modèle de Brennan-Schwartz ou Cox-Ingersol-Ross

C'est un modèle utilisé principalement pour la courbe des taux ([9] page 188) :

$$dX_t = \alpha(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t, t \in [0, T], X_0 > 0.$$

Ici, $\theta = (\alpha, \beta, \sigma) \in \mathbb{R}^2 \times \mathbb{R}^+$, et $b(\theta, x) = \alpha(\beta - x)$, $\sigma(\theta, x) = \sigma\sqrt{x}$. Cette EDS admet une solution forte pour toutes les valeurs du paramètres car b et σ sont localement lipschitziennes. De fait, le calcul stochastique montre que l'on peut choisir les paramètres α et β tels que X_t et Y_i sont presque sûrement strictement positifs. La discrétisation donne la chaîne de Markov :

$$Y_{i+1} = Y_i + \alpha(\beta - Y_i)h + \sigma\sqrt{Y_i}\Delta_i W, Y_0 > 0.$$

Exercice : Le noyau de transition est la densité de la loi $\mathcal{N}(x + \alpha(\beta - x)h, \sigma^2 xh)$. D'où la vraisemblance

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2 h y_{i-1}}} \exp\left[-\frac{(y_i - y_{i-1} - h\alpha(\beta - y_{i-1}))^2}{2\sigma^2 h y_{i-1}}\right], y_0 > 0,$$

et le contraste :

$$C(\theta; y_1, \dots, y_N) = \frac{1}{2}N \log \sigma^2 + \frac{1}{2\sigma^2 h} \sum_{i=1}^N \frac{[y_i - y_{i-1} - h\alpha(\beta - y_{i-1})]^2}{y_{i-1}}.$$

On pose $a = \sigma^2$ et on cherche à minimiser (ce qui ne dépend pas de σ)

$$F : (\alpha, \beta) \mapsto \sum_{i=1}^N \frac{[y_i - y_{i-1} - h\alpha(\beta - y_{i-1})]^2}{y_{i-1}}.$$

Pour tout β , cette fonction est convexe en α :

$$(12) \quad F'_\alpha = -\sum_{i=1}^N \frac{h(\beta - y_{i-1})(y_i - y_{i-1} - h\alpha(\beta - y_{i-1}))}{y_{i-1}}$$

$$(13) \quad F''_\alpha = \sum_{i=1}^N \frac{h^2(\beta - y_{i-1})^2}{y_{i-1}} \geq 0$$

Pour tout α , elle est aussi convexe en β :

$$(14) \quad F'_\beta = -\sum_{i=1}^N \frac{\alpha h(y_i - y_{i-1} - h\alpha(\beta - y_{i-1}))}{y_{i-1}}$$

$$(15) \quad F''_\beta = \sum_{i=1}^N \frac{\alpha^2 h^2}{y_{i-1}} \geq 0$$

et sous réserve de vérifier à l'aide de la matrice D^2F que le 0 du gradient réalise bien un minimum de F , l'étude des estimateurs de minimum de contraste (EMC) donne à résoudre le système :

$$(16) \quad \sum_{i=1}^N \frac{(y_i - y_{i-1})(\beta - y_{i-1}) - \alpha h(\beta - y_{i-1})^2}{y_{i-1}} = 0,$$

$$(17) \quad \sum_{i=1}^N \frac{(y_i - y_{i-1}) - \alpha h(\beta - y_{i-1})}{y_{i-1}} = 0,$$

ce qui donne tout calcul fait :

$$(18) \quad \tilde{\alpha} = \frac{(y_N - y_0)B_N - NA_N}{h(N^2 - D_N B_N)},$$

$$(19) \quad \tilde{b} = \frac{N(y_N - y_0) - D_N A_N}{B_N(y_N - y_0) - NA_N}.$$

en ayant posé :

$$A_N = \sum_{i=1}^N \frac{y_i - y_{i-1}}{y_{i-1}} ; B_N = \sum_{i=1}^N \frac{1}{y_{i-1}} ; D_N = \sum_{i=1}^N y_i.$$

En effet, de la deuxième équation on tire

$$\alpha = \frac{\sum_i \frac{y_i - y_{i-1}}{y_{i-1}}}{h[\beta \sum_i \frac{1}{y_{i-1}} - N]} = \frac{A_N}{h(\beta B_N - N)}$$

et de la première

$$\alpha = \frac{\beta \sum_i \frac{y_i - y_{i-1}}{y_{i-1}} - \sum_i (y_i - y_{i-1})}{h(\beta^2 \sum_i \frac{1}{y_{i-1}} - 2\beta N + \sum_i y_i)} = \frac{\beta A_N - (y_N - y_0)}{h(\beta^2 B_N - 2N\beta + D_N)}.$$

On égalise les deux expressions dont on tire la valeur de $\tilde{\beta}$ que l'on reporte dans l'une des expressions de α pour avoir $\tilde{\alpha}$. Ceci se fait sans connaissance de la variance.

Le contraste en fonction de cette variance est de même nature que les précédents dont le minimum est atteint pour:

$$\tilde{\alpha} = \frac{1}{T} \sum_{i=1}^N \frac{[(y_i - y_{i-1}) - \tilde{\alpha} h(\tilde{\beta} - y_{i-1})]^2}{y_{i-1}}.$$

4.3 Résultats généraux pour l'asymptotique

Ce paragraphe rassemble quelques extraits du chapitre 3 de [7].

Dans ces modèles statistiques paramétrés dominés, une notion importante est celle de l'information de KULLBACK.

Définition 4.3 Si $f_\theta(x)$ est la densité du modèle paramétré par rapport à la mesure dominante, et si α et θ sont deux éléments de l'ensemble Θ tels que $\log \frac{f_\theta}{f_\alpha} \in L^1(\mathbb{P}_\theta)$, alors l'information de KULLBACK $K(\mathbb{P}_\theta, \mathbb{P}_\alpha) = E_\theta(\log \frac{f_\theta}{f_\alpha})$.

Remarquons que $E_\theta(\log \frac{f_\theta}{f_\alpha}) = E_\alpha(\frac{f_\theta}{f_\alpha} \log \frac{f_\theta}{f_\alpha} + 1 - \frac{f_\theta}{f_\alpha}) \geq 0$ puisqu'une étude rapide de la fonction $x \mapsto x \log x + 1 - x$ montre qu'elle est positive ; de plus elle atteint son min en $x = 1$ soit, pour ce qui concerne l'information de KULLBACK, si θ est le vrai paramètre, $a \mapsto K(\mathbb{P}_\theta, \mathbb{P}_a)$ est minimum en $a = \theta$ et vaut alors 0.

Un résultat asymptotique intéressant est le suivant :

Proposition 4.4 Si $\log \frac{f_\alpha}{f_\theta} \in L^1(\mathbb{P}_\theta)$, alors on a la convergence \mathbb{P}_θ -presque sûre lorsque n tend vers l'infini :

$$\frac{1}{n} \sum_{i=1}^n \log \frac{f_\alpha}{f_\theta}(X_i) \rightarrow E_\theta(\log \frac{f_\alpha}{f_\theta}) = -K(\mathbb{P}_\theta, \mathbb{P}_\alpha) \leq 0$$

et maximum en $\alpha = \theta$, où (X_i) est un n -échantillon de la loi \mathbb{P}_θ .

C'est une conséquence de la loi des grands nombres et ce résultat justifie l'idée de prendre pour estimateur la valeur du paramètre qui maximise la log-vraisemblance.

On donne maintenant un exemple où l'on obtient un résultat asymptotique analogue sans avoir l'indépendance. Il s'agit d'une suite stationnaire gaussienne centrée de densité spectrale f (cf cours de séries chronologiques : par exemple un AR(1), c'est à dire en quelque sorte un processus de Ornstein-Ulhenbeck). On sait dans ce cas que la densité d'un vecteur gaussien est donnée par

$$P_f(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n |T_n(f)|}} \exp\left(-\frac{\tilde{x} T_n(f)^{-1} x}{2}\right)$$

où $T_n(f)$ est la matrice de covariance ($\gamma(i-j) = cov(X_i, X_j)$, $1 \leq i, j \leq n$) et $|T_n(f)|$ son déterminant. Dans ce cas, deux résultats de séries chronologiques montre la convergence (1.3.23 et 24,[7] page 30) : $\forall \alpha$,

$$\frac{1}{n} \log |T_n(f_\alpha)| \rightarrow \frac{1}{2\pi} \int_0^{2\pi} \log |2\pi f_\alpha(\lambda)| d\lambda.$$

Par ailleurs, si l'on effectue un changement de base où la matrice T_n est diagonale, on se ramène à une somme de variables aléatoires indépendantes et il vient la convergence \mathbb{P}_θ presque sûre :

$$\frac{1}{n} \tilde{X}_n((T_n(f_\alpha))^{-1} - (T_n(f_\theta))^{-1})X_n - E_\theta\left[\frac{1}{n} \tilde{X}_n((T_n(f_\alpha))^{-1} - (T_n(f_\theta))^{-1})X_n\right] \rightarrow 0$$

Or cette espérance vaut $\frac{1}{n} \text{trace}(T_n(f_\alpha)(T_n(f_\theta))^{-1} - I)$ de limite quand n tend vers l'infini

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{f_\alpha}{f_\theta}(\lambda) d\lambda - 1$$

d'après le lemme 2.5.28 page 71 [7].

Si maintenant l'on prend le log du quotient des deux densités \mathbb{P}_α et \mathbb{P}_θ , en rassemblant ces deux limites, il vient comme pour un n -échantillon :

$$\frac{1}{n} \log \frac{\mathbb{P}_\theta}{\mathbb{P}_\alpha}(x_1, \dots, x_n) \rightarrow -K(f_\theta, f_\alpha)$$

mais qui est ici $\frac{1}{4\pi} \int_0^{2\pi} (\log \frac{f_\alpha}{f_\theta} + \frac{f_\alpha}{f_\theta} - 1)(\lambda) d\lambda$. dans $L^2(\mathbb{P}_\theta)$ ([7] page 91, ex. 3).

4.3.1 Contrastes

De façon plus générale que dans le schéma d'Euler, on a la définition suivante ([7] page 92) :

Définition 4.5 Une fonction de contraste relativement à un paramètre θ_0 est une application K de Θ dans \mathbb{R}^+ admettant un minimum strict en θ_0 .

Un exemple simple de fonction de contraste est l'information de Kullback.

Définition 4.6 Lorsque les expériences sont décrites sur une espace $(\Omega, \mathcal{A}, (\mathcal{F}_t, t \in T), \mathbb{P})$, un **contraste** relatif à un couple (paramètre θ , fonction K) est une application U (indépendante de θ) de $\Theta \times T \times \Omega$ dans \mathbb{R}^+ telle que

(i) $\forall (\alpha, t) \in \Theta \times T, U(\alpha, t, \cdot)$ est \mathcal{F}_t -mesurable.

(ii) $U(\alpha, t, \cdot)$ converge en \mathbb{P}_θ probabilité vers $K(\theta, \alpha)$ lorsque t tend vers l'infini.

Alors, $\hat{\theta}_t = \arg \min(U(\alpha, t), \alpha \in \Theta)$ s'appelle un **estimateur de minimum de contraste** (EMC, noté $C(\theta, K)$).

Dans l'exemple des suites stationnaires gaussiennes centrées de densité spectrale f , la filtration est discrète, (X_1, \dots, X_n) est de densité spectrale f_α , on propose

$$U(\alpha, n, \cdot) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} [\log f_\alpha(\lambda) + \frac{I_n(\lambda)}{f_\alpha(\lambda)}] d\lambda$$

où $I_n(\lambda) = \frac{1}{2\pi n} |\sum_{p=1}^n X_p e^{-ip\lambda}|^2$ est le spectrogramme, estimateur de la densité spectrale.

Exercice : montrer qu'il s'agit bien d'un contraste en précisant la fonction de contraste K .

4.3.2 Consistance des EMC

Définition 4.7 Un estimateur \hat{g}_t de $g(\theta)$, ($\theta \in \Theta$) est **consistant** au point θ si \hat{g}_t converge en \mathbb{P}_θ probabilité vers $g(\theta)$ lorsque t tend vers l'infini.

Il est consistant s'il est consistant en tout point.

Il est fortement consistant si la convergence est presque sûre.

Sur ces notions on a le résultat suivant ([7] page 94) :

Théorème 4.8 On suppose

(i) Θ est un compact de \mathbb{R}^k , $\alpha \mapsto U(t, \alpha)$ et $K(\theta, \alpha)$ sont continues sur Θ .

(ii) On note pour $\varepsilon > 0$, $w(t, \varepsilon) = \sup\{|U(t, \alpha) - U(t, \beta)|, |\alpha - \beta| \leq \varepsilon\}$; il existe une suite ε_k décroissant vers 0 telle que

$$\forall k, \lim_{t \rightarrow \infty} \mathbb{P}_\theta(w(t, 1/k) \geq \varepsilon_k) = 0.$$

Alors, tout EMC est consistant en θ .

Ces hypothèses sont vérifiées dès que $U(t, \cdot)$ est Lipschitzienne sur tout compact avec une constante C_t telle que $\lim_{t \rightarrow \infty} C_t = a$: presque sûrement

$$|U(t, \alpha) - U(t, \beta)| \leq C_t |\alpha - \beta| ;$$

alors on obtient que $w(t, \varepsilon) \leq C_t \varepsilon$ et il suffit de prendre $\varepsilon_k = \frac{1}{\sqrt{k}}$ pour obtenir

$$\mathbb{P}_\theta\{w(t, 1/k) \geq \varepsilon_k\} \leq \mathbb{P}_\theta\{C_t \geq \sqrt{k}\}$$

qui tend vers 0 quand t tend vers l'infini puisque $\lim_{t \rightarrow \infty} C_t = a$.

4.3.3 Vitesse de convergence

([7] page 96)

Il s'agit de trouver par exemple une fonction du temps $v(t)$ telle que

$$\sup_t (v(t))^2 E_\theta[|\hat{\theta}_t - \theta|^2] < \infty,$$

ou bien

$$v(t)[\hat{\theta}_t - \theta] \rightarrow G_\theta$$

en loi, avec G_θ de loi connue ce qui permet d'obtenir des régions de confiance. Par exemple, on recherche des normalités asymptotiques.

Théorème 4.9 (3.3.11 page 98 [7])

Soit (T_n) une suite de vecteurs aléatoires de \mathbb{R}^k tels que (T_n) converge en probabilité vers m et $\sqrt{n}(T_n - m)$ converge en loi vers une gaussienne centrée de matrice de covariance Γ .

Soit g une fonction sur un voisinage V de m dans \mathbb{R}^k de classe C^2 , D^2g bornées, et la matrice jacobienne notée J_g . Alors quand n tend vers l'infini, on a la convergence en loi :

$$\sqrt{n}(g(T_n) - g(m)) \rightarrow \mathcal{N}(0, J_g(m)\Gamma(J_g(m))^t).$$

4.3.4 Lois limites

Une autre notion importante est celle-ci :

Définition 4.10 Si $L_t(\theta)$ est la vraisemblance d'un modèle au temps t , on appelle l'information de FISHER de θ au temps t la matrice $E_\theta[\nabla \log L_t(\theta) \nabla \log \tilde{L}_t(\theta)] = -E_\theta[\partial_{ij}^2 \log L_t(\theta)]$.

On dit qu'un modèle est régulier si

- (i) $\alpha \mapsto L_t(\alpha)$ est de classe C^2 au voisinage de θ ,
- (ii) $\nabla \log L_t(\theta)$ est une variable aléatoire centrée de carré intégrable sous \mathbb{P}_θ ,
- (iii) et $E_\theta[\nabla \log L_t(\theta) \nabla \log \tilde{L}_t(\theta)]$ information de Fisher notée $I_t(\theta)$ est inversible.

Théorème 4.11 ([7] page 101)

Soit un modèle régulier en θ de vraisemblance L et d'information de Fisher I . S'il existe un voisinage U de θ dans Θ et une fonction $h \in L^1(\mathbb{P}_\theta)$ telle que $\forall x, \forall \alpha \in U, \forall i, j, |D_{ij}^2 \log L(\alpha, x)| \leq h(x)$.

Soit un n -échantillon de ce modèle régulier qui donne lieu à un EMV $\hat{\theta}_n$ qui converge vers θ en probabilité. Alors, quand n tend vers l'infini,

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow \mathcal{N}(0, I^{-1}(\theta))$$

en loi sous \mathbb{P}_θ , et on a la convergence en \mathbb{P}_θ -probabilité quand n tend vers l'infini :

$$I(\theta)\sqrt{n}(\hat{\theta}_n - \theta) - \frac{1}{\sqrt{n}}\nabla I_n(\theta) \rightarrow 0.$$

4.4 Comportement asymptotique

4.4.1 Le pas h_n tend vers 0, et $T = nh_n$ tend vers l'infini

La motivation est d'obtenir des estimateurs explicites construits avec les observations $(X_{ih_n}, i = 1, \dots, n)$. La méthode est de construire des contrastes pour remplacer la log-vraisemblance comme on l'a vu plus haut.

(i) Estimation de la dérive

On considère ici que le coefficient de diffusion (la volatilité) est connu et on s'intéresse au modèle

$$dX_t = b(c, X_t)dt + \sigma dW_t$$

en dimension 1. Le schéma d'Euler donne ici

$$\tilde{X}_{ih} - \tilde{X}_{(i-1)h} = b(c, \tilde{X}_{(i-1)h})h + \sigma\sqrt{h}\varepsilon_i$$

où ε_i sont des variables gaussiennes centrées réduites indépendantes. Le contraste est la log-vraisemblance du schéma d'Euler que l'on a déjà donné :

$$c \mapsto C(c; y_1, \dots, y_n) = -\frac{1}{2}n \log \sigma^2 - \frac{1}{2\sigma^2 h} \sum_{i=1}^n [y_i - y_{i-1} - hb(c, y_{i-1})]^2.$$

Sans avoir explicitement l'EMC, en utilisant les résultats du paragraphe précédent, on a quand même des résultats asymptotiques.

Proposition 4.12 *Si \tilde{c}_n est l'estimateur de maximum de contraste, alors on a la convergence en loi lorsque nh_n^3 tend vers 0 :*

$$\sqrt{nh_n}(\tilde{c}_n - c) \rightarrow \mathcal{N}(0, I^{-1}(\theta_0))$$

où $I^{-1}(\theta_0)$ est l'information de Fisher du modèle.

(ii) Estimation simultanée de la dérive et de la volatilité

On complique un peu le modèle en :

$$dX_t = b(c, X_t)dt + a(\sigma, X_t)dW_t, X_0 = x \in \mathbb{R}.$$

On pose $\theta = (c, \sigma)$ et on obtient le contraste :

$$C(\theta; y_1, \dots, y_n) = -\frac{1}{2} \left[\sum_{i=1}^n \log a^2(\sigma, y_{i-1}) + \sum_{i=1}^n \frac{[y_i - y_{i-1} - hb(c, y_{i-1})]^2}{2a(\sigma, y_{i-1})} \right].$$

L'EMC obtenu pour θ n'est pas explicite. La méthode proposée est d'effectuer des développements limités au voisinage de $h_n = 0$.

4.4.2 Le pas $h_n = T/n$ tend vers 0, n tend vers l'infini et T est fixe

Dans ce cadre asymptotique, on a des résultats essentiellement pour la volatilité, fondés sur les propriétés des variations quadratiques. (référence : Donahl, Journal of Applied Probability, 24, 105-114, 1987)

Le fondement de l'étude est le modèle très simple où $dX_t = \sigma dW_t$, auquel cas il est bien connu que l'EMV de σ est la variation quadratique observée ($a = \sigma^2$) :

$$\tilde{a} = \frac{1}{T} \sum_{i=1}^N (x_i - x_{i-1})^2$$

dont on sait que c'est une variable aléatoire de loi a fois un chi-deux à N degrés de liberté ; donc c'est un estimateur sans biais et de plus par le théorème de la limite centrale, on a la convergence en loi lorsque N tend vers l'infini :

$$\sqrt{N}(\tilde{a} - a) \rightarrow \mathcal{N}(0, 2a^2).$$

La généralisation du modèle (en dimension 1) repose sur le développement limité de la densité de transition du modèle discrétisé au voisinage de $h = 0$. (référence : Dacunha-Castelle et Duflo, 1983)

$$dX_t = b(\theta, X_t)dt + \sigma(\theta, X_t)dW_t.$$

Lorsque le paramètre θ est réel, on peut montrer (Genon-Catalot et Jacod 1993, Donahl 1998, Gobet 2000) :

Proposition 4.13 *Lorsque N tend vers l'infini, on a la convergence en loi*

$$\sqrt{N}(\tilde{\theta} - \theta) \rightarrow \frac{Z}{\sqrt{I(\theta)}}$$

où Z et $I(\theta)$ sont deux variables indépendantes, $Z \sim \mathcal{N}(0, 1)$ et $I(\theta) = \frac{2}{T} \int_0^T \left(\frac{a'(\theta, X_s)}{a(\theta, X_s)} \right)^2 ds$.

En multidimensionnel (cf Genon-Catalot et Jacod 1993), lorsque la dérive est nulle, le schéma d'Euler est

$$\tilde{X}_{ih} - \tilde{X}_{(i-1)h} = \sigma(\theta, (i-1)h, X_{(i-1)h})\sqrt{h}\varepsilon_i$$

où ε_i sont des variables gaussiennes centrées réduites indépendantes. L'EMC est explicite et vérifie une convergence en loi vers une loi de mélange en variance de gaussiennes.

4.4.3 Le pas h est fixe, N et $T = Nh$ tendent vers l'infini

On considère une diffusion en dimension 1 autonome (??) et on fait une hypothèse d'ergodicité.

Dacunha-Castelle et Florens-Zmirou (1986) obtiennent un EMV exact de θ avec une convergence en loi qui vient des propriétés des chaînes de Markov ergodiques sur le modèle :

$$dX_t = b(\theta, X_t)dt + \sigma(\theta, X_t)dW_t, X_0 = x \in \mathbb{R}.$$

L'idée sous-jacente est de mesurer la perte d'information induite par la discrétisation par rapport au cas de l'estimation de θ fondée sur l'observation de toute la trajectoire. D'où la proposition de développement asymptotique au voisinage de $h = 0$. Les résultats sont les suivants

. Log-vraisemblance du paramètre $c = (\theta, \sigma)$:

$$l_n(c) = \sum_{i=1}^N p_h(X_{ih}, X_{(i-1)h}, c)$$

. On a la convergence en loi des EMV lorsque N tend vers l'infini :

$$\sqrt{N}(\tilde{c} - c) \rightarrow \mathcal{N}(O, V)$$

où V est une matrice de variance fonction de c . Au voisinage de $h = 0$, cette matrice est de la forme :

$$V_{11} = \frac{1}{h} \int \frac{(b'_\theta(\theta, u))^2}{\sigma^2} \pi_c(du) + C + O(h) ; V_{12} = C + O(h) ; V_{22} = 2\sigma^4 + O(h).$$

L'outil essentiel est une fois de plus le développement du contraste pour h petit. On pose $G_\theta(x)$ primitive de $b(\theta, x$, et $g_c(x) = \frac{b^2(\theta, x)}{\sigma^2} + b'_x(\theta, x)$, alors

$$p_h(x, y; c) = \frac{1}{2\pi\sigma^2 h} \exp\left[-\frac{(x-y)^2}{2\sigma^2 h} + \frac{G_\theta(x) - G_\theta(y)}{\sigma^2}\right] \times E\left[\exp\left(-\frac{h}{2} \int_0^1 h_c(ux + (1-u)y + \sqrt{\sigma^2 h} B_u) du\right)\right]$$

où B est le pont brownien entre 0 et 1.

4.5 Diffusion élastique

Il s'agit du modèle

$$dX_t = (aX_t + b)dt + \sigma X_t^\psi dW_t, \psi \in]0, 1[, t \in [0, T], X_0 \neq 0.$$

Ce modèle peut être intéressant pour modéliser non seulement des courbes de taux, mais aussi des actifs dont la volatilité est influencée par le prix. Ici, $\theta = (\alpha, \beta, \sigma) \in \mathbb{R}^2 \times \mathbb{R}^+$.

La chaîne de Markov associée est solution de l'équation de récurrence

$$x_i = \alpha(\beta - x_{i-1})h + \sigma x_{i-1}^\psi \Delta_i W, i = 1, \dots, N.$$

La densité du noyau de transition est

$$p(x, y; \theta) = \frac{1}{\sqrt{2\pi\sigma^2 x_{i-1}^{2\psi}}} \exp\left[-\frac{(x_i - \alpha(\beta - x_{i-1})h)^2}{2\sigma^2 x_{i-1}^{2\psi}}\right]$$

d'où la vraisemblance de la chaîne de Markov :

$$L(\theta; x_i) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2 x_{i-1}^{2\psi}}} \exp\left[-\frac{(x_i - \alpha(\beta - x_{i-1})h)^2}{2\sigma^2 h x_{i-1}^{2\psi}}\right]$$

et le contraste

$$C(\theta; x_i) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2 h} \sum_{i=1}^N \frac{(x_i - \alpha(\beta - x_{i-1})h)^2}{x_{i-1}^{2\psi}}.$$

L'étude sera analogue à la précédente avec au dénominateur des $x_{i-1}^{2\psi}$ au lieu de x_{i-1} . La concavité de C est de même nature et les estimateurs de maximum de contraste de ce modèle sont solution du système

$$(20) \quad \sum_{i=1}^N \frac{(y_i - y_{i-1})(b - y_{i-1}) - \alpha h (b - y_{i-1})^2}{y_{i-1}^{2\psi}} = 0,$$

$$(21) \quad \sum_{i=1}^N \frac{(y_i - y_{i-1}) - \alpha h (b - y_{i-1})}{y_{i-1}^{2\psi}} = 0,$$

à résoudre en exercice....y compris l'étude de l'asymptotique....

References

- [1] L. BACHELIER : “Théorie de la spéculation” : Annales scientifiques de l’école normale supérieure, 17, 21-88, Paris, 1900.
- [2] O.E. BARNDORFF-NIELSEN and N. SHEPHARD : “Aggregation and model construction for volatility models”, Working paper series n. 10, Center for Analytical Finance, university of Aarhus, 1998.
- [3] R. BELLMAN : “Introduction to the Mathematical Theory of Control Process”, Vol. I, II, Academic Press, New York and London, 1967.
- [4] B.M. BIBBY and M. SØRENSEN : “Martingale estimations functions for discretely observed diffusion processes” : Bernoulli, 1, 17-39, 1995.
- [5] F. BLACK and M. SCHOLES : “The pricing of options and corporate liabilities”, Journal of Political Economy, 3, 637-654, 1973.
- [6] J.C. COX and S.A. ROSS : “ The valuation of options for alternative stochastic processes”, Journal of Financial Economics, 7, 229-263, 1979.
- [7] D. DACUNHA-CASTELLE et M. DUFLO, *Probabilités et Statistique*, tome 2, Masson, Paris, 1983.
- [8] D. DACUNHA-CASTELLE et D. FLORENS-ZMIROU : “Estimation of the coefficient of a diffusion from discrete observations”, Stochastics, 19, 263-284, 1986.
- [9] Rose Ann DANA et Monique JEANBLANC, *Marchés financiers en temps continu, valorisation et équilibre*, Economica, deuxième édition, Paris, 1998.
- [10] G. DEMANGE et J.C. ROCHET : “Méthodes mathématiques de la finance”, Economica.
- [11] M.U. DOTHAN : “Prices in financial Markets”, Oxford University Press, Oxford, 1990.
- [12] V. GENON-CATALOT and J. JACOD : “On the estimation of the diffusion coefficient for multidimensional diffusion processes”, Annales de l’IHP, Proba-Stat, 29, 119-151, 1993.
- [13] V. GENON-CATALOT et D. PICARD, *Eléments de statistique asymptotique*, Mathématiques et Applications, Springer-Verlag, New-York , 1993.
- [14] V. GENON-CATALOT, T. JEANTHEAU and C. LAREDO : “Parameter estimation for discretely observed volatility models”, Bernoulli, 5(5), 1999, 855-872.
- [15] D. GOYEAU et A. TARAZI, *La Bourse*, collection Repères, ed. La découverte, Paris, 2001.
- [16] J.M. HARRISON and D.M. KREPS : “Martingales and arbitrage in multiperiod securities markets”, Journal of Economic Theory, 20, 381-408, 1979.
- [17] J.M. HARRISON and S. PLISKA : “Martingales and stochastic integrals in the theory of continuous trading”, Stochastic Processes and their Applications, 11, 215-260, 1981.

- [18] C. HUANG, “Information structure and equilibrium asset prices”, *Journal of Economic Theory*, 35, 33-71, 1985.
- [19] IBRAGIMOV and R.Z. HAS’MINSKII, *Statistical Estimation, Asymptotic Theory*, Springer-Verlag, New York, 1981.
- [20] I. KARATZAS and S.E. SHREVE : “Brownian Motion and Stochastic calculus”, Springer-Verlag, New York, 1988.
- [21] I. KARATZAS and S.E. SHREVE, *Methods of Mathematical Finance*, Springer-Verlag, New York, 1998.
- [22] M. KESSLER, “Simple and explicit estimating functions for a discretely observed diffusion process”, Research Report 336, Dept theor. Statis., University of Aarhus, 1996.
- [23] M. KESSLER, “Estimation of an ergodic diffusion from discrete observations”, *Scandinavian Journal of Statistics*, 24(2), 211-229, 1997.
- [24] Y. KUTOYANTS, *Parameter Estimation for Stochastic processes*, Heldermann, Berlin, 1984.
- [25] D. LAMBERTON et B. LAPEYRE, *Introduction au calcul stochastique appliqué à la finance*, Ellipses, Paris, 1991.
- [26] S. R. PLISKA, *Introduction to Mathematical Finance*, Blackwell, Oxford, 1998.
- [27] P. PROTTER : “Stochastic Integration and Differential Equations”,
- [28] F. QUITTARD-PINON, *Marchés de capitaux et théorie financière*, Economica, Paris, 1993.
- [29] A.N. SHIRYAYEV, *Essential of Stochastic Finance*, World Scientific, Singapore, 1999.