

TRISTAN LORINO

PROCESSUS STOCHASTIQUES

Février 2005



« ON APPELLE ÇA, un peu obscurément, la loi des grands nombres. Par quoi l'on peut dire à peu près que, si un homme se tue pour telle raison et un autre pour telle autre, dès qu'on a affaire à un très grand nombre, le caractère arbitraire et personnel de ces motifs disparaît, et il ne demeure... précisément, qu'est-ce qui demeure? Voilà ce que j'aimerais vous entendre dire. Ce qui reste, en effet, vous le voyez vous-même, c'est ce que nous autres profanes appelons tout bonnement la moyenne, c'est-à-dire quelque chose dont on ne sait absolument pas ce que c'est. Permettez-moi d'ajouter que l'on a tenté d'expliquer logiquement cette loi des grands nombres en la considérant comme une sorte d'évidence. On a prétendu, au contraire, que cette régularité dans des phénomènes qu'aucune causalité ne régit ne pouvait s'expliquer dans le cadre de la pensée traditionnelle; sans parler de mainte autre analyse, on a aussi défendu l'idée qu'il ne s'agissait pas seulement d'événements isolés, mais de lois, encore inconnues, régissant la totalité. Je ne veux pas vous ennuyer avec les détails, d'autant que je ne les ai plus présents à l'esprit, mais personnellement, il m'importerait beaucoup de savoir s'il faut chercher là-dedans quelque mystérieuse loi de la totalité ou si tout simplement, par une ironie de la Nature, l'exceptionnel provient de ce qu'il ne se produit rien d'exceptionnel, et si le sens ultime du monde peut être découvert en faisant la moyenne de tout ce qui n'a pas de sens! L'une ou l'autre de ces deux conceptions ne devrait-elle pas avoir une influence décisive sur notre sentiment de la vie? Quoi qu'il en soit, en effet, la possibilité d'une vie ordonnée repose toute entière sur cette loi des grands nombres; si cette loi de compensation n'existait pas, il y aurait des années où il ne se produirait rien, et d'autres où plus rien ne serait sûr; les famines alterneraient avec l'abondance, les enfants seraient en défaut ou en excès et l'humanité voletterait de côté et d'autre entre ses possibilités célestes et ses possibilités infernales comme les petits oiseaux quand on s'approche de leur cage. »

MUSIL, *L'homme sans qualités*.

Sommaire

I	SÉRIES CHRONOLOGIQUES	7
1	Introduction	8
1.1	Mesures spectrales — processus ARMA	10
1.2	Prédiction linéaire	14
1.3	Prédiction sur le passé infini	15
1.4	Estimation	17
1.4.1	Moyenne	17
1.4.2	Covariance	18
1.4.3	Modélisation par les MA	19
1.5	Théorème spectral et applications	21
1.6	Équations ARMA canoniques	26
1.7	Covariance et auto-corrélation des ARMA	27
2	Modélisation	30
2.1	Introduction	30
2.2	Modélisations AR et MA — Estimations préliminaires	31
2.2.1	AR	31
2.2.2	MA	32
2.3	Estimation efficace	35
2.4	Processus ARIMA	38
2.5	Modèles multiplicatifs	41
2.6	Envoi	42
2.6.1	Critères de choix	42
2.6.2	Tests d'ajustement	42
3	Modèles autorégressifs non linéaires	44
3.1	Rappels sur les modèles autorégressifs linéaires	44
3.1.1	Cadre univarié	44
3.1.2	Cadre multivarié	45
3.1.3	Retour au cadre univarié	46
3.2	Modèles autorégressifs non linéaires lipschitziens	47
3.2.1	Modèles hétéroscédastiques	48
3.2.2	Modèle autorégressif non linéaire à coefficients aléatoires	49
3.3	Ergodicité	49
3.4	Chaînes de Markov et stabilité	52
3.5	Modèles ARCH et GARCH	54
3.6	Modèles de diffusions limites des modèles GARCH	56
II	THÉORIE DE MARKOV	58

4	Introduction	59
5	Ergodicité	66
6	Entropie	69
III	PROCESSUS STOCHASTIQUES	71
7	Généralités	72
7.1	Espaces gaussiens	73
7.2	Mouvement brownien	73
7.3	Principe d'invariance	74
7.4	Propriétés du brownien	74
7.4.1	Variation quadratique	74
7.4.2	Martingales	75
7.4.3	Théorème d'arrêt — Inégalité de Doob	76
7.4.4	Intégrale de Wiener	78
7.4.5	Équation de Langevin	79
8	Calcul stochastique	80
8.1	Intégrale stochastique d'Ito	80
8.1.1	Filtration	80
8.1.2	Fonctions en escalier	81
8.1.3	Densité des fonctions en escaliers dans $M^2(\mathbb{R}^+)$	82
8.1.4	Intégrale stochastique	82
8.2	L'intégrale stochastique comme martingale	82
8.3	Formule d'Ito	83
8.3.1	Introduction	83
8.3.2	Formule générale	83
8.3.3	Localisation	84
8.3.4	Cas vectoriel	85
8.3.5	Intégration par parties	87
8.4	Formule de Girsanov	87
8.4.1	Formule de Cameron-Martin	87
8.4.2	Théorème de Girsanov	87
8.4.3	Critères	88
9	Processus de comptage	89
9.1	Rappels concernant les martingales	89
9.2	Processus à variation prévisible	89
9.3	Processus de comptage	90
9.3.1	Cas univarié	90
9.3.2	Cas multivarié	91
9.4	Théorème de la limite centrale	92
9.5	Résidus	92
9.6	Théorie du produit intégral (ou produit infini)	93
9.7	Approche markovienne des processus de comptage	94
IV	ÉQUATIONS DIFFÉRENTIELLES STOCHASTIQUES	96

10 Introduction	97
10.1 Existence et unicité de solutions fortes	97
10.2 Exemples	97
10.3 Solutions faibles d'EDS	98
10.3.1 Modèle de Cox – Ingersoll – Ross	99
10.3.2 Absolue continuité de la loi de diffusions sous changement de dérive	99
10.3.3 Équations linéaires	101
10.3.4 Autre EDS	101
10.3.5 Pont brownien	102
11 Propriétés des EDS	103
11.1 Caractère markovien des EDS	103
11.1.1 Propriété de Markov des solutions des EDS	103
11.1.2 Générateurs et EDS	106
11.2 Équations différentielles stochastiques rétrogrades	108
11.3 Lien avec les EDP semi-linéaires	110
11.3.1 Rappel sur la formule de feynman-Kac	110
11.3.2 Généralisation de la formule de Feynman-Kac	113
11.4 Applications des EDSR aux solutions de viscosité d'une famille d'EDP non linéaires du second ordre	113
11.4.1 Équation projective (<i>forward</i>)	113
11.4.2 Équation rétrograde (<i>backward</i>)	114
12 Statistique des diffusions	116
12.1 Introduction	116
12.2 Processus d'Ornstein-Uhlenbeck	118
12.3 Markov et les diffusions	119
12.3.1 Étude des estimateurs du maximum de vraisemblance	124
12.4 Estimateurs empiriques	126
V MODÈLE LINÉAIRE GÉNÉRALISÉ	129
13 Introduction	130
13.1 Modèle linéaire classique	130
13.2 Modèle linéaire général	131
13.2.1 Estimation par les moindres carrés ordinaires	131
13.2.2 Estimation par les moindres carrés pondérés	131
13.2.3 Estimation par le maximum de vraisemblance sous l'hypothèse de normalité	132
13.2.4 Estimation robuste des écarts-types	135
14 Modèle linéaire généralisé	136
14.1 Présentation	136
14.1.1 Les équations de vraisemblance	138
14.1.2 Algorithmes	139
14.1.3 Simplification lors de l'utilisation d'un lien canonique	141
14.1.4 Ajustement	142
14.1.5 Étude des résidus	142
14.2 Données binaires	142
14.2.1 Méthode itérative de Newton-Raphson	144
14.2.2 Méthode du scoring de Fisher	144
14.3 Modèle linéaire généralisé à effets mixtes	144

14.3.1	Définition	144
14.3.2	Estimation des paramètres	145
VI	ÉQUATIONS D'ESTIMATION GÉNÉRALISÉES	147
15	Quasi-vraisemblance	148
15.1	Vraisemblance marginale	148
15.2	Vraisemblance conditionnelle	148
15.3	Quasi-vraisemblance	148
15.4	Méthode de Newton-Raphson	150
15.5	Méthode de Fisher	150
15.6	Conditions d'application	151
16	Équations d'estimation généralisées	152
16.1	Modèle	152
16.2	Estimation des paramètres	153
16.2.1	Estimation de β^*	153
16.2.2	Estimations de α et ϕ	153
16.2.3	Estimation de la variance de $\hat{\beta}^*$	154
16.3	Différentes matrices de travail $R(\alpha)$	154
16.4	Extensions des GEE	156

Première partie

SÉRIES CHRONOLOGIQUES

1

Introduction

Définition 1.1 — On appelle **processus** une suite $(X_n)_n$ de variables aléatoires.

Définition 1.2 — Un processus $(X_t)_t$ à valeurs réelles ou complexes est dit **du second ordre** si

$$\mathbb{E}[|X_t|^2] < \infty \quad \forall t \in T.$$

Pour un processus de second ordre, on appelle **covariance** la fonction Γ définie sur $T \times T$ par

$$\Gamma(s, t) = \mathbb{E}[(X_s - \mathbb{E}(X_s))(X_t - \mathbb{E}(X_t))].$$

Remarques — Il est utile de noter que :

1. $\mathcal{L}^2(\mathbb{P}) \subset \mathcal{L}^1(\mathbb{P})$;
2. si X est réelle, alors Γ est réelle, symétrique et semi-définie positive : $\forall t_1, \dots, t_n \in T, \forall a \in \mathbb{C}^n$,

$$\sum_{i,j} a_i \Gamma(t_i, t_j) \bar{a}_j \geq 0 ;$$

3. Γ semi-définie positive au sens complexe $\Rightarrow \Gamma$ hermitienne, i.e. $\Gamma(s, t) = \overline{\Gamma(t, s)}$;
4. Γ réelle et semi-définie positive $\Rightarrow \Gamma$ symétrique, i.e. $\Gamma(s, t) = \Gamma(t, s)$.

Théorème 1.1 — Si Γ est une fonction réelle, symétrique et semi-définie positive sur $T \times T$, il existe un processus $(X_t)_{t \in T}$ réel du second ordre de covariance Γ .

Théorème 1.2 — Si Γ est une fonction réelle, symétrique et semi-définie positive sur $T \times T$, il existe un processus $(X_t)_{t \in T}$ gaussien centré de covariance Γ .

Définition 1.3 — Un processus réel $(X_t)_t$ est dit **gaussien** si $\forall n, \forall (t_1, \dots, t_n) \subset T$, la variable aléatoire vectorielle $(X_{t_1}, \dots, X_{t_n})$ est gaussienne. Un processus gaussien est du second ordre.

Définition 1.4 — Un processus du second ordre est dit **centré** si $\mathbb{E}(X_t) = 0$, $\forall t \in T$.

Définition 1.5 — Un processus du second ordre $(X_n)_n$ est dit **stationnaire au sens large** si la moyenne $\mathbb{E}(X_n)$ est constante et si la covariance $\Gamma(n, m)$ ne dépend que de la différence $n - m$, i.e. $\exists \gamma : \mathbb{Z} \rightarrow \mathbb{C}$ telle que

$$\begin{aligned}\Gamma(n, m) &= \gamma(n - m) \\ \gamma(n) &= \mathbb{E}\left[(X_n - \mathbb{E}(X_n))\overline{(X_0 - \mathbb{E}(X_0))}\right] \\ \gamma(n) &= \mathbb{E}(X_n \overline{X_0}) \quad \text{si les variables sont centrées} \\ \mathbb{E}(X_n) &= \mathbb{E}(X_0) \quad \forall n.\end{aligned}$$

Notation — On notera **SLC** un processus du second ordre stationnaire au sens large et centré.

Définition 1.6 — Le **coefficient de corrélation** $\rho(n)$ est une mesure de la dépendance entre l'instant 0 et l'instant n :

$$\rho(n) = \frac{\gamma(n)}{\gamma(0)}.$$

Si ρ est proche de 1, la mémoire est dite « longue ».

Remarque — γ est une fonction semi-définie positive telle que :

1. $\gamma(0)$ est réel, positif ou nul ;
2. $\gamma(n) = \overline{\gamma(-n)}$;
3. $|\gamma(n)| \leq \gamma(0)$, $\forall n$.

Si le SLC est réel, alors γ est réelle et $\gamma(n) = \gamma(-n)$.

Proposition 1.1 — Si $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ est paire et semi-définie positive, alors il existe un SLC gaussien de fonction de covariance γ .

Remarque — Le processus gaussien qui vient d'être construit a la propriété suivante :

$$\mathbb{P}(X_{n_1} \in A_1, \dots, X_{n_k} \in A_k) = \mathbb{P}(X_{n_1+n} \in A_1, \dots, X_{n_k+n} \in A_k).$$

Définition 1.7 — Un processus $(X_n)_n$ est dit **stationnaire au sens strict** si $\forall k, n_1, \dots, n_k, n$,

$$\mathbb{P}(X_{n_1} \in A_1, \dots, X_{n_k} \in A_k) = \mathbb{P}(X_{n_1+n} \in A_1, \dots, X_{n_k+n} \in A_k).$$

Théorème 1.3 (Ergodicité) — Soit $f(X_0, \dots, X_d) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Alors

$$\frac{1}{N} \sum_{i=1}^N f(X_i, \dots, X_{i+d}) \xrightarrow{p.s.} \mathbb{E}(f(X_0, \dots, X_d)).$$

1.1 Mesures spectrales — processus ARMA

Soit $\Pi = \mathbb{R}/2\pi\mathbb{Z}$ le **tore**. On note $e_1(t) = e^{it}$, $e_n(t) = e^{int}$ et $\bar{e}_n(t) = e^{-int} = e_{-n}(t)$. La mesure de Lebesgue est invariante par translation :

$$\int_{\Pi} f(x) dx = \int_{\Pi} f(x+u) dx, \quad \forall u \in \Pi.$$

$(e_n)_n$ est un système orthonormé dans $\mathcal{L}^2(\lambda)$, où λ est la mesure de Lebesgue.

$$\begin{aligned} \|e_n\|_2^2 &= \int_{\Pi} |e_n|^2 d\lambda \\ &= \int_{\Pi} e_n \bar{e}_n d\lambda \\ &= \int_{\Pi} 1 d\lambda \\ \langle e_n, e_m \rangle &= \int_{\Pi} e_n \bar{e}_m d\lambda \\ &= \int_{\Pi} e_{n-m} d\lambda \\ &= \begin{cases} 0 & \text{si } n \neq m \\ 1 & \text{si } n = m. \end{cases} \end{aligned}$$

$\sum_{-\infty}^{+\infty} a_n e_n$ converge dans $\mathcal{L}^2(\lambda)$ ssi $\sum_{-\infty}^{+\infty} |a_n|^2 < \infty$.

$$\begin{aligned} l^2(\mathbb{C}) &\longrightarrow \mathcal{L}_{\mathbb{C}}^2(\Pi) \\ (a_n)_n \in &\longmapsto \sum_{-\infty}^{+\infty} a_n e_n : \quad \text{isométrie bijective} \end{aligned}$$

$$\begin{aligned} \left\| \sum a_n e_n \right\|_2^2 &= \langle \sum a_n e_n, \sum a_n e_n \rangle \\ &= \sum_n \sum_m a_n \bar{a}_m \langle e_n, e_m \rangle \\ &= \sum_{-\infty}^{+\infty} |a_n|^2. \end{aligned}$$

Rappel — Une isométrie est toujours injective ; ici, elle est de plus surjective.

$\{e_n\}_n$ est une base orthonormale de $\mathcal{L}^2(\lambda)$: pour $f \in \mathcal{L}^2(\lambda)$,

$$f = \sum_{-\infty}^{+\infty} a_n e_n,$$

avec $a_n = \int f e_{-n} d\lambda$, qui est le n^e coefficient de Fourier de f (noté $\hat{f}(n)$). Nous avons

$$\langle f, e_n \rangle = \langle \sum_{-\infty}^{+\infty} a_k e_k, e_n \rangle = a_n.$$

$f \mapsto (\hat{f}(n))_{n \in \mathbb{Z}}$ est la **transformation de Fourier**. C'est une isométrie de $\mathcal{L}_{\mathbb{C}}^2(\Pi)$ sur $l_{\mathbb{C}}^2(\mathbb{Z})$. Si μ est une mesure bornée sur Π , sa transformée de Fourier est la fonction $\mathcal{F}\mu$ ou $\hat{\mu}$ définie sur \mathbb{Z} par

$$\mathcal{F}\mu(n) = \int_{\Pi} e_n \, d\mu .$$

Propriété 1.1 — *L'application $\mu \mapsto \hat{\mu}$ est injective.*

Théorème 1.4 (Herglotz) — *La fonction γ sur \mathbb{Z} est une covariance si et seulement si il existe une mesure positive μ sur Π telle que*

$$\gamma(n) = \int_{\Pi} e_n \, d\mu .$$

Définition 1.8 — *Si X est un SLC de covariance γ_X , la mesure μ_X telle que*

$$\gamma_X(n) = \int e_n \, d\mu_X$$

*s'appelle la **mesure spectrale** de X . De plus, si μ_X a une densité par rapport à λ , i.e. $\mu_X = f_X \cdot \lambda$ (pour $f_X \in \mathcal{L}_+^1(\lambda)$), cette densité s'appelle la **densité spectrale**.*

Rappel — (μ_n) converge étroitement vers μ si et seulement si $\forall f \in \mathbb{C}_b$,

$$\int f \, d\mu_n \longrightarrow \int f \, d\mu .$$

Proposition 1.2 — *Si (X_n^k) est une suite de SLC telle que, pour tout n , $\lim_{k \rightarrow \infty} X_n^k$ existe dans \mathcal{L}^2 , si on appelle X_n cette limite, alors le processus $X = (X_n)$ est un SLC et μ_X est la limite étroite des $\mu_{X_n^k}$.*

Proposition 1.3 — *Si $\sum_{-\infty}^{+\infty} |\gamma_X(n)|^2 < \infty$, alors μ_X est la mesure ayant pour densité la fonction*

$$\sum_{-\infty}^{+\infty} |\gamma_X(k) e_{-k}| .$$

De même si $\sum_{-\infty}^{+\infty} |\gamma_X(n)| < \infty$ (et dans ce cas, $\sum \gamma_X(k) e_{-k}$ est continue).

Définition 1.9 — *On appelle **bruit blanc** de variance σ^2 une suite de v.a. réelles, centrées, appartenant à \mathcal{L}^2 , de variance σ^2 et 2 à 2 non corrélées. On note $(U_n) \in BB(\sigma^2)$.*

Nota — Un bruit blanc est un SLC.

Proposition 1.4 — Un SLC U est un $BB(\sigma^2)$ si et seulement si $\mu_U = \sigma^2\lambda$.

Proposition 1.5 — Si X est un SLC et si $a \in l^1(\mathbb{Z})$, alors le processus Y défini par

$$Y_n = \sum_{-\infty}^{+\infty} a_k X_{n-k}$$

est un SLC dont la mesure spectrale μ_Y est donnée par

$$\mu_Y = \left| \sum_{-\infty}^{+\infty} a_k e^{-k} \right|^2 \mu_X .$$

Définition 1.10 — L'opération de passage de X à Y s'appelle une **opération de filtrage**. On dit que Y est la transformée de X par le **filtre de fonction de transfert** $f = \sum a_k e^{-k}$.

Proposition 1.6 — Si U est un $BB(\sigma^2)$ et si $a \in l^2(\mathbb{Z})$, alors

$$Y_n = \sum_{-\infty}^{+\infty} a_k U_{n-k}$$

définit un SLC de mesure spectrale

$$\left| \sum_{-\infty}^{+\infty} a_k e^{-k} \right|^2 \cdot \sigma^2 \lambda .$$

Définition 1.11 (MA) — Si dans la proposition précédente, on suppose $a_k = 0$ pour $k < 0$, le processus Y obtenu s'appelle un **MA**(∞)¹. Si de plus $a_k = 0$ pour $k > q$, le processus Y s'appelle un **MA**(q).

Exemple — Le processus

$$\begin{aligned} X_n &= U_n + \rho U_{n-1} + \cdots + \rho^k U_{n-k} + \cdots \\ &= U_n + \rho(U_{n-1} + \rho U_{n-2} + \cdots) \\ &= U_n + \rho X_{n-1} \end{aligned}$$

est un **processus autorégressif d'ordre 1**.

Proposition 1.7 — Si $|\rho| \neq 1$ et si U est un SLC ou un BB , il existe un SLC X tel que

$$X_n - \rho X_{n-1} = U_n, \quad \forall n .$$

1. MA pour *moving average.*, c.-à-d. *moyenne mobile*.

Définition 1.12 (ARMA) — On appelle **processus ARMA** d'ordre (p, q) un SLC réel X satisfaisant à une équation du type

$$a_0X_n + a_1X_{n-1} + \cdots + a_pX_{n-p} = b_0U_n + b_1U_{n-1} + \cdots + b_qU_{n-q},$$

avec $U \in BB$ et $a_i, b_j \in \mathbb{R}$.

Moving average MA(q)	$X_n = \sum_{k=0}^q b_k U_{n-k} .$
Auto-regressive AR(p)	$U_n = \sum_{k=0}^p b_k X_{n-k} .$
ARMA (p, q)	$a_0X_n + a_1X_{n-1} + \cdots + a_pX_{n-p} = b_0U_n + b_1U_{n-1} + \cdots + b_qU_{n-q} .$

On note

$$P(z) = a_0 + a_1z + \cdots + a_pz^p$$

et

$$Q(z) = a_0 + a_1z + \cdots + a_qz^q .$$

Définition 1.13 — Soit B l'opérateur de retard (*shift*) :

$$(BX)_n = X_{n-1} \text{ et } (B^k X)_n = X_{n-k} .$$

La définition revient donc à

$$P(B) \cdot X = Q(B) \cdot U .$$

Cette équation est appelée **équation ARMA**.

Théorème 1.5 — Si Z est un SLC et P un polynôme n'ayant pas de racines de module 1, alors il existe un SLC X tel que $P(B)X = Z$.

Corollaire 1.1 — Si P est un polynôme n'ayant pas de racines de module 1, il existe des ARMA(p, q), i.e. des SLC réels X tels que $P(B)X = Q(B)U$.

Proposition 1.8 — Si X est un ARMA solution de $P(B)X = Q(B)U$ où $U \in BB(\sigma^2)$, alors X a une densité spectrale égale à

$$\frac{|Q \circ e_{-1}|^2}{|P \circ e_{-1}|^2} \sigma^2 .$$

Définition 1.14 — On dit qu'un ARMA est un **processus à spectre rationnel**.

1.2 Prédiction linéaire

Soient Y, X_1, \dots, X_n dans \mathcal{L}^2 . On cherche $\hat{Y} = b + a_1 X_1 + \dots + a_n X_n$ tel que $\mathbb{E}[(Y - \hat{Y})^2]$ soit minimum.

Théorème 1.6 — *La meilleure prédiction de Y par une fonction affine $X = (X_1, \dots, X_n)^t$ est donnée par*

$$\hat{Y} = \mathbb{E}(Y) + \sum_{i=1}^n a_i (X_i - \mathbb{E}(X_i)) ,$$

i.e.

$$\hat{Y} = \mathbb{E}(Y) + a^t (X - \mathbb{E}(X))$$

avec a racine du système

$$\Gamma_X a = (\text{cov}(X_i, Y)) .$$

De plus, si Γ_X est inversible, on a

$$\text{Var}(Y - \hat{Y}) = \text{Var}(Y) - a^t \Gamma_X a ,$$

où $\text{Var}(Y - \hat{Y})$ est la **variance résiduelle**.

Proposition 1.9 (Algorithme de Dubin-Levinson) — *Soit X un SLC. On suppose que $\gamma_X(0) > 0$. Soit $\hat{X}_{n+1} = \Phi_{n1} X_n + \Phi_{n2} X_{n-1} + \dots + \Phi_{nn} X_1$ la meilleure prédiction de X_{n+1} en fonction de X_n, X_{n-1}, \dots . Soit $v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$ l'erreur de prédiction. Alors les Φ_{nj} et les v_n sont données par les 3 équations de récurrence suivantes :*

$$\begin{cases} \Phi_{nn} = \frac{\gamma(n) - \sum_{j=1}^{n-1} \Phi_{n-1,j} \gamma(n-j)}{v_{n-1}} , \\ \Phi_{nj} = \Phi_{n-1,j} - \Phi_{nn} \Phi_{n-1,n-j} , \\ v_n = v_{n-1} (1 - \Phi_{nn}^2) , \end{cases}$$

avec les conditions initiales $\Phi_{11} = \gamma(1)/\gamma(0)$ et $v_0 = \gamma(0)$.

Définition 1.15 — *On appelle **fonction d'autocorrélation partielle** la fonction $r(n)$ définie par*

$$r(n) = \phi_{n,n}, \quad n \geq 1 .$$

Remarque — $r(n)$ grand pour n grand : « mémoire longue ».

Proposition 1.10 —

$$\begin{aligned} r(n) &= \text{corr} (X_{n+1} - \text{proj}_{\mathcal{H}}(X_{n+1}), X_1 - \text{proj}_{\mathcal{H}}(X_1)) \\ &= \frac{\langle X_{n+1} - \text{proj}_{\mathcal{H}}(X_{n+1}), X_1 - \text{proj}_{\mathcal{H}}(X_1) \rangle}{\|X_1 - \text{proj}_{\mathcal{H}}(X_1)\| \cdot \|X_{n+1} - \text{proj}_{\mathcal{H}}(X_{n+1})\|} , \end{aligned}$$

avec $\mathcal{H} = \text{ev}(X_2, \dots, X_n)$.

Remarque —

$$\text{ev}(X_1, \dots, X_n) = \mathcal{H} \oplus \mathbb{R}(X_1 - \text{proj}_{\mathcal{H}}(X_1)) .$$

Remarque — La connaissance de $r(n)$ entraîne celle des $\gamma(n)/\gamma(0)$.

Proposition 1.11 (Algorithme de l'innovation) — Soit X un SLC. $(X_n - \hat{X}_n)$ est une suite de v.a. deux à deux non corrélées. Par conséquent, $(X_k - \hat{X}_k)_{k=1, \dots, n}$ constituent une base de l'espace vectoriel $\text{ev}(X_1, \dots, X_n)$ avec la convention $\hat{X}_1 = 0$. On pose

$$\hat{X}_{n+1} = \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) .$$

Alors les $\theta_{n,j}$ et les v_n sont données par les 3 équations de récurrence suivantes :

$$\begin{cases} v_0 = \Gamma(1,1) , \\ \theta_{n,n-k} = \frac{\Gamma(n+1,k+1) - \sum_{j=0}^{k-1} (\theta_{k,k-j} - \theta_{n,n-j} \theta_{n,n-j}) v_j}{v_k} & k = 0, 1, \dots, n-1 , \\ v_n = \Gamma(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j . \end{cases}$$

1.3 Prédiction sur le passé infini

Soit X un processus du second ordre centré. Soient

$$\begin{aligned} H_n^X &= \overline{\text{ev}(X_i, i \neq n)} \\ &= \left\{ \text{limites dans } \mathcal{L}^2 \text{ de combinaisons linéaires des } X_i = \sum_{\text{finie}} a_k X_{n-k} \right\} . \end{aligned}$$

H_n^X est une suite croissante ($H_n^X \subset H_{n+1}^X$).

Notations — On note

$$\begin{aligned} H_{-\infty}^X &= \bigcap_n H_n^X \\ H_{\infty}^X &= \overline{\bigcup_n H_n^X} . \end{aligned}$$

Définition 1.16 — $H_{-\infty}^X$ est appelé le « *passé infini* ».

Proposition 1.12 — On a

$$H_{-\infty}^X \subseteq H_n^X \subseteq H_{n+1}^X \subseteq H_{\infty}^X .$$

Définition 1.17 — Un processus du second ordre est dit *singulier* si

$$H_{\infty}^X = H_{-\infty}^X .$$

Définition 1.18 — Un processus du second ordre est dit **régulier** si

$$H_{-\infty}^X = \{0\}.$$

Remarque — Un processus à la fois singulier et régulier est identiquement nul.

Lemme 1.1 — Si H est un espace de Hilbert et H_n une suite croissante (respectivement décroissante) de sous-espaces fermés de H , alors pour tout x de H , la suite de projections $\text{proj}_{H_n}(x)$ converge vers $\text{proj}_{H_\infty}(x)$, où $H_\infty = \overline{\cup H_n}$ (resp. $H_\infty = \cap H_n$).

Théorème 1.7 (Décomposition de Wold) — Si X est un processus du second ordre centré, il existe deux processus X^r et X^s , respectivement régulier et singulier, orthogonaux entre eux et tels que

$$X_n = X_n^r + X_n^s$$

i.e.

$$H_n^X = H_n^{X^r} \oplus H_n^{X^s}.$$

Cette décomposition est unique. De plus,

$$H_{-\infty}^X = H_{-\infty}^{X^s}.$$

Notation — On note $p_n = \text{proj}_{H_n^X}$.

Remarque —

$$X_n^s = p_{-\infty}(X_n)$$

Proposition 1.13 — Si X est un SLC, il existe une isométrie B de H^X sur lui-même telle que

$$(BX)_n = X_{n-1}.$$

De plus,

$$Bp_n = p_{n-1}B.$$

Remarque —

$$p_n(X) = \lim_{p \rightarrow \infty} p_{(X_n, X_{n-1}, \dots, X_{n-p})}(X).$$

Proposition 1.14 — Soit X un SLC. Alors le processus

$$U_n = X_n - p_{n-1}(X_n)$$

est un BB non nul ssi X n'est pas singulier. De plus, on a

$$p_\infty B = Bp_\infty,$$

$$\sigma_U^2 = \lim_{n \rightarrow \infty} \frac{\det \Gamma_{n+1}}{\det \Gamma_n}$$

et

$$BU_n = U_{n-1}.$$

Définition 1.19 — *Ce processus s'appelle l'innovation de X .*

Corollaire 1.2 — *Si X est un SLC, les parties singulière et régulière de sa décomposition de Wold sont des SLC.*

Proposition 1.15 — *Les 3 énoncés suivants sont équivalents :*

1. X est régulier ;
2. il existe un BB U tel que $H_n^X = H_n^U$, $\forall n$ (U est l'innovation) ;
3. il existe un BB W et une suite $c \in l^2(\mathbb{N})$ telle que¹

$$X_n = \sum_0^{\infty} c_k W_{n-k} .$$

Remarque Cette proposition signifie « l'identité » entre les processus réguliers et les MA(∞).

Proposition 1.16 — *Si W est un BB tel que $H_n^W = H_n^X \forall n$, alors il existe des scalaires λ_n tels que $(|\lambda_n|)_n$ soit une suite constante et*

$$w_n = \lambda_n U_n ,$$

où U est l'innovation de X .

Remarque — Ceci signifie l'unicité « essentielle » du bruit blanc telle que voulue en (2) de l'avant-dernière proposition — cependant qu'il n'y a pas unicité en (3).

Proposition 1.17 — *Un SLC X est un MA(q) ssi $\gamma_X(n) = 0$ dès que $|n| > q$.*

Remarque — Si X est un processus gaussien, l'innovation est une suite de v.a. gaussiennes indépendantes, centrées et de même variance.

1.4 Estimation

1.4.1 Moyenne

Soit x_1, \dots, x_n une série expérimentale qui est une réalisation de X_1, \dots, X_n , processus stationnaire large que l'on notera X .

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

est un estimateur (sans biais) de la moyenne m . Est-il convergent ?

1. W n'est pas forcément l'innovation — mais on peut prendre l'innovation pour W .

Proposition 1.18 —

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = \mu_X(\{0\}) .$$

Théorème 1.8 — Si $\mu_X(\{0\}) = 0$, alors \bar{X}_n converge vers m en moyenne quadratique (i.e. dans \mathcal{L}^2). De plus, si X a une densité f_X continue en 0, alors \bar{X}_n converge p.s. vers m .

Théorème 1.9 — Si

$$X_n = b + \sum_{j=-\infty}^{+\infty} a_j Z_{n-j} ,$$

où $a \in l^1$ et $(Z_n)_n$ est une suite de v.a. i.i.d. centrées et de variance σ^2 , et si

$$\alpha = \sum_{-\infty}^{+\infty} a_j \neq a$$

alors

$$\sqrt{n}\bar{X}_n \xrightarrow{\mathcal{L}} \mathcal{N}(b, \alpha^2 \sigma^2) .$$

1.4.2 Covariance

Soit N la longueur de la série. On suppose qu'elle est centrée. On désire estimer la covariance. Soit $\forall k \geq 0$,

$$\tilde{\gamma}_X = \frac{1}{N-k} \sum_{i=1}^{N-k} X_i X_{i+k} .$$

Il est sans biais. Cependant un problème demeure : la fonction $\tilde{\gamma}$ ainsi définie n'est pas nécessairement semi-définie positive.

Définition 1.20 — On définit

$$\hat{\gamma} = \frac{1}{N} \sum_{i=1}^{N-k} X_i X_{i+k} ,$$

avec $N \geq 50$ et $k \leq \frac{N}{4}$.

Proposition 1.19 — $\hat{\gamma}$ est un estimateur asymptotiquement sans biais.

Définition 1.21 — On définit

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} .$$

Théorème 1.10 — Si X est un SLC gaussien tel que $\sum_{k \in \mathbb{Z}} \gamma(k) < \infty$, alors :

1. $\hat{\gamma}(k)$ est un estimateur p.s. convergent de $\gamma(k)$;
2. on a $\forall K \in \mathbb{N}$,

$$\sqrt{N}(\hat{\gamma}_N(i) - \gamma(i))_{i=0, \dots, K} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma) ,$$

où

$$\Gamma_{ij} = \sum_{m \in \mathbb{Z}} (\gamma(m) \cdot \gamma(m+i+j) + \gamma(m) \cdot \gamma(m+i-j)) .$$

De plus

$$\lim_{N \rightarrow \infty} N \mathbb{E} \left[(\hat{\gamma}_N(i) - \gamma(i)) (\hat{\gamma}_N(j) - \gamma(j)) \right] = \Gamma_{ij} .$$

Remarque —

$$\sum_{k \in \mathbb{Z}} |k| \gamma(k) < \infty \iff \gamma \in l^2(\mathbb{Z}) .$$

Théorème 1.11 — Sous les mêmes hypothèses que précédemment, on a que

$$(\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(k)) \xrightarrow{\mathcal{L}} \mathcal{N} \left((\rho(1), \dots, \rho(k)) , \frac{1}{n} W \right) ,$$

où la **formule de Bartlett** donne

$$W_{ij} = \sum_{k=-\infty}^{+\infty} \left\{ \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho(k)^2 - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i)^2 \right\} .$$

Entre autre,

$$W_{ii} = \sum_{k=1}^{+\infty} \{ \rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k) \}^2 .$$

1.4.3 Modélisation par les MA

MA(q) : $\gamma(i) = 0$ si $i > q$. Pour $i > q$,

$$\begin{aligned} W_{ii} &= 1 + 2\rho(1)^2 + \dots + 2\rho(q)^2 \\ &= \Delta_{ii} . \end{aligned}$$

Proposition 1.20 —

$$\sqrt{N}\hat{\rho}(i) \rightsquigarrow \mathcal{N}(0, \Delta_{ii}) .$$

Remarque — On peut faire le test de l'hypothèse H_0 : le MA est d'ordre inférieur ou égal à q . Pour le niveau $\alpha = 0,05$, on rejette l'hypothèse si

$$|\hat{\rho}(i)| > \frac{1,96\sqrt{\Delta}}{\sqrt{N}}$$

pour un $i > q$.

Remarque — Modéliser un MA, c'est déterminer $q, b_0, \dots, b_q, \sigma_U^2$; pour ce faire :

1. on choisira le premier q pour lequel on ne rejette pas l'hypothèse de base ;
2. on résoudra ensuite le système

$$\hat{\gamma}(k) = \sigma_U^2 \sum_{l=k}^q b_l b_{l-k} .$$

Définition 1.22 — Si X est un SLC, on appelle **périodogramme** la fonction aléatoire

$$I_N = \frac{1}{N} \left| \sum_{k=1}^N X_k e^{-k} \right|^2 .$$

Proposition 1.21 —

$$I_N = \sum_{|k| < N} \hat{\gamma}(k) e^{-k} .$$

Remarque — Dans les *bons* cas,

$$f_X = \sum_{-\infty}^{+\infty} \gamma(k) e^{-k} ,$$

d'où I_N apparaît comme étant un **estimateur empirique** de f_X .

Théorème 1.12 — Les mesures $(I_N \lambda)$ convergent p.s. étroitement vers μ_X lorsque X est stationnaire strict et ergodique.

Théorème 1.13 — Si X est un SLC gaussien et si $\sum |k| \cdot |\gamma(k)| < \infty$ pour toute fonction borélienne bornée Φ à valeur dans \mathbb{R}^d , on a

$$\lim \mathbb{E}[I_N(\Phi)] = I(\Phi)$$

et

$$\sqrt{N}(I_N(\phi) - I(\Phi)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma(\Phi)) ,$$

avec

$$\Gamma(\Phi) = \int \Phi \Phi^t f_X^2 d\lambda .$$

1.5 Théorème spectral et applications

Soit $f : \mathbb{Z} \rightarrow \mathbb{C}$, $f \in l^2(\mathbb{Z})$. Alors f s'écrit sous la forme

$$f(n) = \sum_{-\infty}^{+\infty} \hat{f}_k e_{-k}(n).$$

On note

$$\gamma_X(n) = \int e_n d\mu_X.$$

Définition 1.23 — Si (E, \mathcal{E}, μ) est un espace mesuré σ -fini, on appelle **mesure aléatoire** de base μ sur (E, \mathcal{E}) toute isométrie de $\mathcal{L}_{\mathbb{C}}^2(\mu)$ dans $\mathcal{L}_{\mathbb{C}}^2(\Omega, \mathcal{A}, \mathbb{P})$.

Soit $\mathcal{E}_{\mu} = \{A \in \mathcal{E} \mid \mu(A) < \infty\}$. On note Z une mesure aléatoire. Soit $A \in \mathcal{E}_{\mu}$: $\mathbf{1}_A \in \mathcal{L}_{\mathbb{C}}^2(\mu)$. $Z(\mathbf{1}_A) = Z(A)$ est l'image de $\mathbf{1}_A$ par l'isométrie. Z est bien une mesure :

- (i) $Z(\emptyset) = 0$;
- (ii) si $A, B \in \mathcal{E}_{\mu}$ et $A \cap B = \emptyset$, alors

$$\begin{aligned} \mathbb{E}[Z(A)\overline{Z(B)}] &= \int \mathbf{1}_A \overline{\mathbf{1}_B} d\mu \\ &= 0, \end{aligned}$$

i.e. Z est à accroissements orthogonaux ;

- (iii) si $A, B \in \mathcal{E}_{\mu}$ et $A \cap B = \emptyset$, alors

$$Z(A \cup B) = Z(A) + Z(B) ;$$

- (iv) on a

$$\left. \begin{array}{l} (A_n) \in \mathcal{E}_{\mu}, A_n \cap A_m = \emptyset \\ \sum_1^{\infty} \mu(A_n) < \infty \end{array} \right\} \implies \sum_1^{\infty} \mathbf{1}_{A_n} \text{ converge dans } \mathcal{L}^2(\mu) \text{ vers } \mathbf{1}_{\cup A_n}$$

$$Z(\cup A_n) = \sum_1^{\infty} Z(A_n).$$

Définition 1.24 — Si toutes les v.a. $Z(f)$ pour $f \in \mathcal{L}^2(\mu)$ sont centrées, on dira que Z est centrée.

Notation — On note

$$\begin{aligned} Z : \mathcal{L}_{\mathbb{C}}^2(\mu) &\longrightarrow \mathcal{L}_{\mathbb{C}}^2(\Omega, \mathcal{A}, \mathbb{P}) \\ f &\longmapsto Z(f) = \int f dZ = \int f(u) dZ(u) \end{aligned}$$

Réciproquement — Si Z vérifie les points (i), (ii) et (iv), on lui associe une mesure aléatoire : pour $A \in \mathcal{E}_\mu$,

$$Z\left(\sum_{i=1}^N \alpha_i \mathbf{1}_{A_i}\right) = \sum_{i=1}^N \alpha_i Z(A_i).$$

Lemme 1.2 — Si I est une isométrie de $E \subset H$ dans H' , il existe un prolongement unique \bar{I} de I dans \bar{E} . On a

$$\bar{I}(\bar{E}) = \overline{I(E)}.$$

Remarque — La mesure aléatoire est appelée **processus spatial** (ou **champ spectral**).

Proposition 1.22 — Il existe une gaussienne centrée X indexée par \mathcal{E}_μ telle que

$$\mathbb{E}[X(A)X(B)] = \mu(A \cap B).$$

Théorème de Karhunen Si X est un processus du second ordre centré, défini sur $(\Omega, \mathcal{A}, \mathbb{P})$, et s'il existe un espace (E, \mathcal{E}, μ) et une fonction a tels que la covariance de X s'écrive

$$\mathbb{E}[X_s \bar{X}_t] = \int_E a(s, u) \overline{a(t, u)} d\mu(u)$$

(quels que soient t et $a(t, \cdot) \in \mathcal{L}_\mathbb{C}^2(\mu)$), alors il existe une mesure aléatoire Z de base μ telle que

$$X_t = \int_E a(t, u) dZ(u).$$

Remarque — On a

$$Z(a(t, \cdot)) = X_t.$$

Théorème 1.14 (Représentation spectrale) — Si X est un SLC, il existe une mesure aléatoire Z_X de mesure μ_X telle que

$$X_n = \int_\Pi e_n dZ_X.$$

Nota —

$$X_n = Z_X(e_n).$$

Proposition 1.23 — Si μ_X est à support fini, alors il existe des v.a. A_j non corrélées et des $t_j \in \Pi$ tels que

$$X_n = \sum A_j e^{int_j}.$$

Définition 1.25 — U est un processus spatial s'il existe une certaine mesure ν telle que $\forall A, B$ boréliens,

$$\mathbb{E}[U(A)\overline{U(B)}] = \nu(A \cap B).$$

Proposition 1.24 — Soit X un SLC. Les quatre propositions suivantes sont équivalentes :

(i) il existe un BB U et $a \in l^2(\mathbb{Z})$ tels que

$$X_n = \sum_{k=-\infty}^{+\infty} a_k U_{n-k};$$

(ii) il existe $c \in l^2(\mathbb{Z})$ telle que

$$\gamma_X(n-m) = \sum_{k=-\infty}^{+\infty} c_{k-n} \bar{c}_{k-m};$$

(iii) X a une densité;

(iv) X a une densité de la forme

$$\left| \sum_{k=-\infty}^{+\infty} d_k e_{-k} \right|^2$$

avec $d \in l^2$.

Remarque — Il n'y a pas unicité dans (ii) et (iv) (et (i)). La densité spectrale est unique; son écriture, non.

Proposition 1.25 — Si X est un SLC, les trois propriétés suivantes sont équivalentes :

(i) X est régulier;

(ii) X a une densité spectrale de la forme $|\sum_{k=0}^{\infty} a_k e_{-k}|^2$;

(iii) il existe un BB U et $a \in l^2(\mathbb{N})$ tels que

$$X_n = \sum_{k=0}^{+\infty} a_k U_{n-k}.$$

Théorème 1.15 (Szego - Kolmogorov) — Si X est un SLC et σ^2 est la variance de son innovation, si $\mu_X = f_X \cdot \lambda + \nu$ est la décomposition de Lebesgue de μ_X par rapport à λ , alors

$$\sigma^2 = \exp \left(\int_{\Pi} \log f_X d\lambda \right),$$

et si $\sigma^2 > 0$ (i.e. le processus n'est pas singulier), alors $f_X \cdot \lambda$ et ν sont les mesures spectrales respectivement des parties régulière et singulière de la décomposition de Wold de X .

Corollaire 1.3 — Soit X un SLC ayant une densité spectrale f_X . Alors

$$X \text{ régulier} \iff \log(f_X) \in \mathcal{L}^1(\lambda).$$

Théorème 1.16 — Si X est un SLC et $f \in \mathcal{L}^2(\mu_X)$, alors le processus Y défini par

$$Y_n = \int e_n f dZ_X$$

est un SLC de mesure spectrale

$$\mu_Y = |f|^2 \mu_X .$$

Définition 1.26 — On l'appelle l'*image de X par le filtre de réponse f* et on note

$$Y = A_f X .$$

Remarque — On a

$$Z_Y(g) = Z_X(fg)$$

pour tout $g \in \mathcal{L}^2(\mu_Y) = \mathcal{L}^2(|f|^2 \mu_X)$.

Définition 1.27 — Le *filtre passe-bande* est le filtre

$$f = \mathbf{1}_B ,$$

pour un ensemble $B \subset \Pi$.

Définition 1.28 — f est appelée *fonction de transfert*.

Proposition 1.26 — Si U est un BB et $a \in l^2(\mathbb{Z})$, alors le processus

$$X_n = \sum_{-\infty}^{+\infty} a_k U_{n-k}$$

est égal à $A_f U$, où

$$f = \sum_{-\infty}^{+\infty} a_k e_{-k} .$$

On obtient de cette façon tous les processus $A_f U$ pour $f \in \mathcal{L}^2(\mu)$.

Proposition 1.27 — Si X est un SLC, si $f \in \mathcal{L}^2(\mu_X)$ et $g \in \mathcal{L}^2(|f|^2 \mu_X)$, alors $fg \in \mathcal{L}^2(\mu_X)$ et

$$A_{fg}(X) = A_g(A_f(X)) .$$

Lemme 1.3 — Si $Y = A_f X$ et $Z = A_g X$ et si $Y_n = Z_n$ pour un n , alors $f = g$ dans $\mathcal{L}^2(\mu_X)$.

Théorème 1.17 — Si $Y = A_h X$, il existe un filtre A_k tel que $X = A_k Y$ ssi

$$\mu_X(\{h = 0\}) = 0 ,$$

et dans ce cas, $k = 1/h$.

Remarques —

1. $Y_n = \sum_{-\infty}^{+\infty} c_k X_{n-k}$ est la « convolution » sur \mathbb{Z} . Le passage de Y à X la « déconvolution » ;
2. Soit $Y_n = \sum_{-\infty}^{+\infty} c_k X_{n-k}$: si n est le temps, le filtre n'est pas réalisable (car il faut connaître le futur) ;
3. Soit $Y_n = \sum_0^{+\infty} c_k X_{n-k}$: si n est le temps, le filtre est réalisable et qualifié par suite de **causal**.

Notation — On considère un ARMA :

$$\begin{aligned} X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p} &= U_n + b_1 U_{n-1} + \cdots + b_q U_{n-q} \\ P(z) &= 1 + a_1 z + \cdots + a_p z^p \\ Q(z) &= 1 + b_1 z + \cdots + b_q z^q \\ h_p &= P \circ e_{-1} \\ h_q &= Q \circ e_{-1} \\ Ah_p X &= Ah_q U. \end{aligned}$$

On notera

$$A_p X = A_Q U.$$

Théorème 1.18 — *Si P n'a pas de racine de module 1, alors il existe un unique SLC X tel que*

$$A_p X = A_Q U.$$

Théorème 1.19 (Fejer - Riesz) — *F est une fraction rationnelle telle que $F \circ e_{-1}$ soit réelle positive et intégrable ssi il existe une fraction rationnelle irréductible Q/P telle que pour tout z de module 1, on ait*

$$F(z) = \left| \frac{Q}{P}(z) \right|^2,$$

avec P sans racine de module 1.

Lemme 1.4 — $\forall u \in \mathbb{C}, u \neq 0$, si $|z| = 1$, alors

$$(z - u)\left(z - \frac{1}{\bar{u}}\right) = -\frac{z}{\bar{u}}|z - u|^2.$$

Remarque — Soit $f : \Pi \rightarrow \mathbb{R}$ définie par

$$f(t) = \left| \frac{Q}{P} \circ e_{-1}(t) \right|^2.$$

Cette écriture n'est pas unique. On peut multiplier Q/P par :

- (i) des constantes de module 1 ;
- (ii) z^m avec $m \in \mathbb{Z}$;

(iii) $|u|^{\frac{z-(1/\bar{u})}{z-u}}$, avec $u \neq 0$.

Définition 1.29 — Un produit de fonctions d'un de ces trois types s'appelle un **produit de Blaschke**.

Proposition 1.28 — Deux fractions rationnelles ont des modules égaux sur $C = \{|z| = 1\}$ ssi leur produit est un produit de Blaschke.

1.6 Équations ARMA canoniques

Rappel — Si $u \neq 0$,

$$|1 - ue_{-1}| = |u|^2 |1 - \frac{1}{\bar{u}} e_{-1}|^2.$$

Lemme 1.5 — Si P n'a pas de racine de module 1, alors

$$\frac{Q}{P}(z) = \sum_{-\infty}^{+\infty} c_k z^k$$

pour z dans une couronne ouverte contenant $\{|z| = 1\}$. De plus, les c_k tendent vers 0 exponentiellement vite lorsque $z \rightarrow \pm\infty$. Enfin, si toutes les racines de P sont de module supérieur à 1, alors $c_k = 0$ pour $k < 0$.

Proposition 1.29 — Si $A_P X = A_Q U$ avec P sans racine de module 1, alors

$$X_n = \sum_{-\infty}^{+\infty} c_k U_{n-k},$$

les $(c_k)_k$ étant ceux du lemme précédent. Si P a toutes ses racines de module supérieur à 1, alors

$$X_n = \sum_{k=0}^{+\infty} c_k U_{n-k},$$

et en particulier X est régulier.

Proposition 1.30 — Si Q a toutes ses racines de module supérieur à 1, alors il existe $(d_k)_k$ tendant exponentiellement vers 0 et telle que

$$U_n = \sum_{k=0}^{+\infty} d_k X_{n-k},$$

et en particulier X est régulier.

Corollaire 1.4 — Si P et Q ont toutes leurs racines de module supérieur à 1, alors U est un multiple de l'innovation de X .

Définition 1.30 — P est dit **sublime** (respectivement **quasi-sublime**) si toutes les racines de P sont de module supérieur à 1 (resp. supérieur ou égal à 1).

Théorème 1.20 — Si X est un ARMA, il existe deux polynômes P et Q et un bruit blanc U tels que :

- (i) $A_P X = A_Q U$;
- (ii) P est sublime et Q quasi-sublime ;
- (iii) P et Q sont premiers entre eux et $P(0) = Q(0) = 1$.

De plus, pour toute relation $A_{P'} X = A_{Q'} W$ satisfaite par X , on a $d^\circ P \leq d^\circ P'$ et $d^\circ Q \leq d^\circ Q'$. Si p et q sont les degrés respectivement de P et Q , on dira que X est de type minimal (p, q) . La relation $A_P X = A_Q U$ avec les propriétés (ii) et (iii) est unique et s'appelle la **relation canonique** de X . Toute équation $A_{P'} X = A_{Q'} W$ avec $d^\circ P = d^\circ P'$ et $d^\circ Q = d^\circ Q'$ est dite de type minimal.

Proposition 1.31 — Si $A_P X = A_Q U$ est la relation canonique de X , alors U est l'innovation de X .

Proposition 1.32 — Un SLC régulier X est un AR(p) ssi $r(n) = 0$ pour $n \geq p$, où r est la fonction d'autocorrélation partielle.

1.7 Covariance et auto-corrélation des ARMA

Un ARMA X peut être défini de trois manières :

- (i) sa covariance $(\gamma_X(n))$;
- (ii) le triplet (a, b, σ^2) (si a et b sont de dimensions resp. p et q , et si $\sigma^2 = \sigma_U^2$, alors il y a $p + q + 1$ paramètres) ;
- (iii) $X_n = \sum_0^\infty c_k U_{n-k}$, soient (c_k) et σ_U^2 .

Passage de (ii) à (iii) — Les (c_k) sont les coefficients du développement en série de Laurent de P/Q :

$$\frac{1 + b_1 z + \cdots + b_q z^q}{1 + a_1 z + \cdots + a_p z^p} = \sum c_k z^k .$$

Passage de (iii) à (i) —

$$\begin{aligned}
 \gamma_X(n) &= \mathbb{E}(X_0, X_n) \\
 &= \mathbb{E}\left[\left(\sum_0^\infty c_k U_{n-k}\right)\left(\sum_0^\infty c_l U_{n-l}\right)\right] \\
 &= \sigma_U^2 \sum_0^\infty c_{l-n} c_l \\
 &= \sigma_U^2 \sum_{l=n}^\infty c_{l-n} c_l .
 \end{aligned}$$

Passage de (i) à (ii) —

$$\begin{aligned}
 X_n + a_1 X_{n-1} + \dots + a_p X_{n-p} &= U_n + b_1 U_{n-1} + \dots + b_q U_{n-q} \\
 \forall s \geq 0, \quad \mathbb{E}[(X_n + \dots) X_{n-s}] &= \mathbb{E}[(U_n + \dots) X_{n-s}] \\
 \text{i.e.}
 \end{aligned}$$

$$\gamma(s) + a_1 \gamma(s-1) + \dots + a_p \gamma(s-p) = \mathbb{E}[U_n X_{n-s}] + b_1 \mathbb{E}[U_{n-1} X_{n-s}] + \dots + b_q \mathbb{E}[U_{n-q} X_{n-s}] ,$$

dites **équations de Yule - Walker** .

Dès que $s > q$,

$$\gamma(s) + a_1 \gamma(s-1) + \dots + a_p \gamma(s-p) = 0$$

(car $U_{n-q} \perp H_{n-s}^X$, $U_{n-q+1} \perp H_{n-s}^X$, etc).

D'où les p équations suivantes :

$$\begin{cases}
 \gamma(q+1) + a_1 \gamma(q) + \dots + a_p \gamma(q+1-p) &= 0 \\
 \gamma(q+2) + a_1 \gamma(q+1) + \dots + a_p \gamma(q+2-p) &= 0 \\
 \vdots & \\
 \gamma(q+p) + a_1 \gamma(q+p-1) + \dots + a_p \gamma(q) &= 0
 \end{cases}$$

Posons

$$R(p,q) = \begin{pmatrix} \gamma(q) & \dots & \gamma(q+1-p) \\ \vdots & & \vdots \\ \gamma(q+p-1) & \dots & \gamma(q) \end{pmatrix} .$$

Alors le système équivaut à

$$R(p,q)a = -r(p,q) ,$$

où $r(p,q) = (\gamma(q+1), \dots, \gamma(q+p))^t$. On admet que $R(p,q)$ est inversible. Alors

$$a = -R^{-1}(p,q) \cdot r(p,q) .$$

Soit $Y_n = \sum_0^p a_k X_{n-k}$, avec $a_0 = 1$. Y est un MA(q).

$$\begin{aligned}
 \gamma_Y(n) &= \sum_{0 \leq k, l \leq p} a_k a_l \gamma_X(n+k-l) \quad (\text{et } \gamma_Y(n) = 0 \text{ si } |n| > q) \\
 &= \sigma_U^2 \sum_{k=n}^q b_k b_{k-n} \quad \text{pour } 0 \leq n \leq q .
 \end{aligned}$$

Nous sommes donc en présence de $q + 1$ équations, qui vont nous permettre de trouver les b_k, b_{k-n} et σ_U^2 .

Passage de (ii) à (i) — a, b, σ_U^2 et l'écriture $\sum c_k U_{n-k}$ permettent d'obtenir, *via* le système de Yule-Walker, $\gamma(0), \dots, \gamma(q)$. γ est solution de l'équation de récurrence liée au polynôme $P(z)$, qui est sublime.

Théorème 1.21 — *La covariance d'un ARMA décroît exponentiellement vite vers 0.*

Attention — On considèrera dorénavant Q sublime et

$$U_n = \sum_{k=0}^{\infty} d_k X_{n-k}$$

avec $|d_k| \searrow 0$ exponentiellement vite.

Proposition 1.33 —

$$\begin{aligned} p_{n-1}(X_n) &= p_{H_{n-1}^X}(X_n) \\ &= - \sum_{k=1}^{\infty} d_k X_{n-k}. \end{aligned}$$

Théorème 1.22 — *La fonction d'autocorrélation partielle d'un ARMA tend exponentiellement vite vers 0.*

Corollaire 1.5 — $\|p_{n-1}(X_n) - p^s(X_n)\|_2$ tend exponentiellement vite vers 0 lorsque $s \rightarrow \infty$ (p^s est la projection sur $ev(X_{n-1}, \dots, X_{n-s})$).

Théorème 1.23 — *Si $A_P X = A_Q U$ est l'équation canonique de X , et $(Q/P)(z) = \sum_0^{\infty} c_k z^k$, alors*

$$p_n(X_{n+j}) = Z_X \left[e_{n+j} \left[1 - \left(\sum_{k=0}^{j-1} c_k e_{-k} \right) \left(\frac{P}{Q}(e_{-1}) \right) \right] \right]$$

et l'erreur de prédiction vaut

$$\sigma^2 \left(\sum_0^{j-1} c_k^2 \right).$$

2

Modélisation

2.1 Introduction

On se donne x_1, \dots, x_N : il s'agit alors de trouver un ARMA(p,q) tel que la série expérimentale soit une représentation des processus

$$A_P X = A_Q U .$$

Nos **objectifs** sont la prédiction, le contrôle et l'étude scientifique. Mais ici, nous ne nous intéresseront qu'à la prédiction. On évoquera la stationnarité. Elle peut laisser apparaître des périodicités, que l'on tâchera d'éliminer. On tentera d'utiliser des modèles linéaires de la forme

$$X_n = \sum_{-\infty}^{+\infty} c_k U_{n-k}$$

et plus particulièrement ceux s'écrivant

$$X_n = \sum_0^{+\infty} c_k U_{n-k} .$$

Un tel processus a une densité f_X . Dans la pratique, on verra essentiellement des MA(q) :

$$\sum_0^q c_k U_{n-k} ,$$

c-à-d qu'on approxime la densité f_X par des polynômes. Or la classe des fractions rationnelles est plus importante que celle des polynômes; d'où l'on utilisera aussi les fractions rationnelles \rightarrow processus ARMA(p,q). On respectera le **principe de parcimonie** : on approximera toujours par un processus ayant le moins de coefficients possible.

Le plan de modélisation se compose de deux étapes :

- la phase d'identification du modèle, qui consiste à déterminer p et q ;
- la phase d'estimation du modèle, qui consiste à déterminer a , b et σ^2 .

On commence par modéliser par un MA(T) — on peut toujours approximer un ARMA par un MA :

$$\text{ARMA} \rightarrow \text{MA}(-\infty) \rightarrow \text{MA}(T) \text{ ou } \text{AR}(S)$$

Puis on essaie de modéliser par un ARMA(p,q) tel que $p \leq s$ et $q \leq T$ (c'est le principe de parcimonie). Enfin, c'est la phase d'estimation.

Nota — Le principe de parcimonie permet d'éviter le **surajustement**, qui survient quand on cherche à ajuster trop parfaitement.

2.2 Modélisations AR et MA — Estimations préliminaires

2.2.1 AR

Soit l'AR(p) avec p connu :

$$X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p} = U_n ,$$

d'où

$$X_n = U_n - (a_1 X_{n-1} + \cdots + a_p X_{n-p}) .$$

On note

$$\tilde{X}_n = -(a_1 X_{n-1} + \cdots + a_p X_{n-p}) .$$

a_i est le coefficient de la régression de X_n sur $(X_{n-1}, \dots, X_{n-p})$.

On note

$$\begin{aligned} \Gamma_p a &= \gamma_p \\ R(p,0)a &= -r(p,0) , \end{aligned}$$

d'où

$$a = -\Gamma_p^{-1} \gamma_p$$

et

$$\sigma^2 = \gamma(0) - a^t \gamma_p .$$

Les **estimateurs empiriques** sont :

$$\begin{aligned} \hat{a} &= -\hat{\Gamma}_p^{-1} \hat{\gamma}_p , \\ \hat{\sigma}^2 &= \hat{\gamma}(0) - \hat{a}^t \hat{\gamma}_p . \end{aligned}$$

Théorème 2.1 — Si X est un $AR(p)$ d'équation canonique $A_p X = U$ où U est i.i.d. de moyenne nulle et de variance σ^2 , alors les estimateurs empiriques satisfont à

$$\sqrt{n}(\hat{a} - a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Gamma_p^{-1})$$

et

$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^2 .$$

Théorème 2.2 — Si U est i.i.d. et si ϕ_l est le vecteur de la régression de X_n sur $(X_{n-1}, \dots, X_{n-p})$, alors pour $l > p$,

$$\sqrt{n}(\hat{a} - \phi_l) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \Gamma_l^{-1})$$

et en particulier,

$$\sqrt{n} \hat{r}(l) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) .$$

Remarque — Soit X_n, \dots, X_{n-l} et $\hat{X} = \phi - 1X_{n-1} + \dots + \phi_l X_{n-l}$. On a que $\phi_l = r(l)$. Si $AR(p)$: pour $p < l$,

$$X_n = -a_1 X_{n-1} - \dots - a_p X_{n-p} + U_n ,$$

d'où $-a_k = \phi_k$, pour $1 \leq k \leq p$.

On a

$$\Gamma_n a = -\gamma_n$$

et

$$\sqrt{n} \hat{r}(l) \rightarrow \mathcal{N}(0, 1)$$

(toujours si $l > p$).

Sur la figure 2.1, les deux lignes pointillées donnent un intervalle de confiance de 95 % pour $\hat{r}(l)$.

Proposition 2.1 — Pour toute covariance, si a est solution de $\Gamma_p a = -\gamma_p$, alors le polynôme $1 + a_1 z + \dots + a_p z^p$ est quasi-sublime.

2.2.2 MA

1^{re} méthode À q fixé, on résout le système d'équations (via le procédé de Newton)

$$\hat{\gamma}(k) = \sum_{l=k}^q b_l b_{l-k} .$$

2^eméthode (théorique) Si $MA(\infty) : f_x = |Q \circ e_{-1}|^2 \sigma^2$. Si $MA(q) : f_x = \sum_{|k| \leq q} \gamma(k) e_{-k}$. D'où

$$\gamma(0) + \sum_{k=1}^q \gamma(k)(z^k + z^{-k}) = \sigma^2 Q(z) Q\left(\frac{1}{z}\right)$$

pour $z = e_{-1}$. On cherche Q . Le membre de gauche s'écrit $R(z + \frac{1}{z})$ où R est un polynôme de degré q :

$$\begin{aligned} \left(z + \frac{1}{z}\right)^k &= z^k + \binom{k}{1} z^{k-1} + \dots \\ &= z^k + \frac{1}{z} + \binom{k}{1} \left(z^{k-2} + \frac{1}{z^{k-2}}\right) + \dots \\ z^q + z^{-q} &= \left(z + \frac{1}{z}\right)^q + \binom{q}{1} \left(z^{q-2} + \frac{1}{z^{q-2}}\right) + \dots \end{aligned}$$

D'où le résultat. Finalement, l'équation est

$$R\left(z + \frac{1}{z}\right) = \sigma^2 Q(z) Q\left(\frac{1}{z}\right).$$

Soient $\alpha_1, \dots, \alpha_q$ les racines de R :

$$Z + \frac{1}{z} = \alpha_i \implies x_i \text{ et } y_i$$

et l'une des deux est de module strictement supérieur à 1 ; disons x_i . Alors

$$\begin{cases} Q(z) = \prod (z - x_i) \\ Q(0) = 1 \end{cases}$$

3^eméthode De la série expérimentale x_1, \dots, x_N on tire $\hat{\gamma}(k)$. On trouve un polynôme sublime tel que $A_P X = U$ (où U est l'innovation), *i.e.* un AR(s) avec s grand. Puis on trouve le MA : $X = A_{1/P} U$. Q est alors le début de la série $1/P$. L'inconvénient de cette méthode est qu'on ne sait pas si le polynôme trouvé est sublime (*i.e.* si on a abouti à la représentation canonique).

4^eméthode elle s'appuie sur la définition suivante.

Définition 2.1 — Si X est un SLC de densité spectrale f continue et positive, on appelle **fonction d'autocorrélation inverse** la fonction définie par

$$\gamma_i(n) = \int_{\Pi} e_n\left(\frac{1}{f}\right) d\lambda.$$

Si X est un MA, alors $f_X = \sigma^2 |Q \circ e_{-1}|^2 \geq 0$ et γ_i est la covariance d'un AR :

$$\frac{1}{f_X} = \sigma^{-2} \frac{1}{|Q \circ e_{-1}|^2}.$$

Reste à trouver un estimateur $\hat{\gamma}_i$ de γ .

- Soit la modélisation AR(T) avec T grand :

$$P(z) = \sum_0^T a_i z^i$$

$$A_P X = U \Rightarrow \frac{\sigma^2}{|P \circ e_{-1}|^2}$$

$$Y = A_P V \text{ avec } V \in BB(\sigma^2) \Rightarrow \frac{|P \circ e_{-1}|^2}{\sigma^2}$$

et la fonction d'autocorrélation de Y est l'inverse de celle de U .

- On prend pour estimation $\hat{\gamma}_i$ de γ_i la covariance estimée de Y :

$$\gamma_Y = \sum a_i a_{i+k}$$

- On fait une modélisation AR sur $\hat{\gamma}_i$.

5^e méthode On fait tourner l'algorithme de l'innovation (même inconvénient qu'avec la 3^e méthode). D'après Yule-Walker : $R(p, q)a = -r(p, q)$.

Proposition 2.2 — Si X est un ARMA, alors $R(p, q)$ est inversible.

On a

$$\begin{aligned} a &= -R(p, q)^{-1} r(p, q) , \\ a_i &= f_i(\gamma(p+q-1), \dots), \quad 1 \leq i \leq p , \\ \hat{a}_i &= f_i(\hat{\gamma}(p+q-1), \dots), \quad 1 \leq i \leq p , \\ \hat{a} &= -\hat{R}(p, q)^{-1} \hat{r}(p, q) . \end{aligned}$$

Proposition 2.3 — Si X est gaussien, l'estimateur \hat{a} est convergent, asymptotiquement normal, mais pas efficace.

Nota — Tester si \hat{P} est sublime ; écrire $A_P X = A_Q U$: $A_{\hat{P}} X$ doit être un MA. Puis $(x_1, \dots, x_N) \rightarrow (y_1, \dots, y_{N-p})$:

$$\begin{cases} x_p + \hat{a}_1 x_{p-1} + \dots + \hat{a}_p x_1 = y_1 \\ x_{p+1} + \dots = y_2 \\ \vdots \end{cases}$$

On fait une modélisation MA $\rightarrow \hat{b}, \hat{\sigma}^2$.

2.3 Estimation efficace

Soient p et q fixés, et

$$\begin{aligned}\zeta_n &= (x_1, \dots, x_n)^t \\ \chi_n &= (X_1, \dots, X_n)^t \\ \Gamma_n &= \mathbb{E}[\chi_n^t \chi_n].\end{aligned}$$

Soit $\theta \in \Theta$; $f(\theta, \zeta_n)$; soit $\tilde{\theta}_n$ tel que

$$f(\tilde{\theta}_n, \zeta_n) = \sup_{\theta \in \Theta} f(\theta, \zeta_n).$$

$$\begin{aligned}\mathbb{P}[\chi_n \in V(\zeta_n)] &= \int_{V(\zeta_n)} f(\theta, \zeta) d\zeta \longrightarrow \int f(\theta, \zeta_n) \\ \text{i.e. } \frac{1}{\lambda(V(\zeta_n))} \int_{V(\zeta_n)} f(\theta, \zeta) d\zeta &\longrightarrow f(\theta, \zeta_n).\end{aligned}$$

On considère un ARMA(p, q) gaussien : $\theta = \{a, b, \sigma^2\}$ avec $\Theta \subseteq \mathbb{R}^{p+q} \times \mathbb{R}^+$ et ouvert. La fonction de vraisemblance est

$$L_n(\theta, \chi_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det \Gamma_n|^{1/n}} \exp\left(-\frac{1}{2} \chi_n^t \Gamma_n^{-1} \chi_n\right),$$

d'où

$$\log L_n(\theta, \chi_n) = -\frac{1}{2} \left(n \log 2\pi + \log(\det \Gamma_n) + \chi_n^t \Gamma_n^{-1} \chi_n \right).$$

Définition 2.2 — On appelle *log-vraisemblance approchée*

$$h_n^1(\theta, \chi_n) = -\frac{1}{2} \left(n \log 2\pi + n \log \sigma^2 + \chi_n^t \Gamma_n^{-1} \chi_n \right)$$

$$\begin{aligned}\sigma^2 &= \lim \frac{\det \Gamma_{n+1}}{\det \Gamma_n}, \\ \log \sigma^2 &= \lim \left(\log (\det \Gamma_{n+1}) - \log (\det \Gamma_n) \right), \\ n \log \sigma^2 &= \log (\det \Gamma_{n+1}).\end{aligned}$$

Proposition 2.4 — Si X est un ARMA, il existe une constante c telle que

$$|n \log \sigma^2 - \log (\det \Gamma_{n+1})| \leq c.$$

Définition 2.3 — L'estimateur du maximum de vraisemblance est $\tilde{\theta}$, tel que

$$h_n^1(\tilde{\theta}_n, \zeta_n) = \sup_{\theta} h_n^1(\theta_n, \zeta_n).$$

Théorème 2.3 — L'estimateur $\tilde{\theta}_n$ est un estimateur convergent, asymptotiquement normal et efficace. De plus,

$$\sqrt{n}(\tilde{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, J(\theta)^{-1}),$$

où

$$J_{kl}(\theta) = \int \frac{\frac{\partial f}{\partial \theta_k} \cdot \frac{\partial f}{\partial \theta_l}}{f(\theta)^2} d\lambda,$$

où f est la densité spectrale.

Proposition 2.5 — Si X est un SLC régulier d'innovation U , et si l'on note \hat{U}_k la régression de U_k sur (X_1, \dots, X_n) , alors

$$\sigma^2 \chi_n^t \Gamma_n^{-1} \chi_n = \sum_{-\infty}^n \hat{U}_k^2,$$

où σ^2 est la variance de U . De plus, cette quantité ne dépend plus de σ^2 .

Proposition 2.6 — Si X est un ARMA(p, q) d'équation canonique (a, b, σ^2) , il existe un BB V ayant même futur que X et tel que

$$X_n + \sum_{k=1}^p a_k X_{n+k} = V_n + \sum_{l=1}^q b_l V_{n+l}$$

i.e.

$$\forall n, \quad \overline{ev(X_n, X_{n+1}, \dots)} = \overline{ev(V_n, V_{n+1}, \dots)}.$$

ARMA

$$\begin{aligned} \sum_{k=0}^p a_k X_{n-k} &= \sum_{l=0}^q b_l U_{n-l} \\ \sum_{k=0}^p a_k X_{n-k} &= \sum_{l=0}^q b_l V_{n-l}, \end{aligned}$$

d'où

$$\begin{aligned} \sum_{k=0}^p a_k \hat{X}_{n-k} &= \sum_{l=0}^q b_l \hat{U}_{n-l} \\ \sum_{k=0}^p a_k \hat{X}_{n-k} &= \sum_{l=0}^q b_l \hat{V}_{n-l} \end{aligned}$$

et

$$\begin{aligned}\widehat{U}_k &= 0 \quad \text{si } k > N \\ \widehat{X}_k &= X_k \quad \text{si } 1 \leq k \leq N \\ \widehat{V}_k &= 0 \quad \text{si } k \leq 0.\end{aligned}$$

— **Étape 1** : calcul des \widehat{V}_k : on se donne un vecteur $\alpha \in \mathbb{R}^q$ et on pose

$$\widehat{V}_{N-p+j} = \alpha_j.$$

$$\begin{cases} n = N - p : & \sum_{k=0}^p \underbrace{a_k \widehat{X}_{N-p+k}}_{\text{connu}} = \widehat{V}_{N-p} + \sum_{l=1}^q \underbrace{b_l \widehat{V}_{N-p+l}}_{\text{connu}} & \Rightarrow & \widehat{V}_{N-p} \\ & \vdots & & \\ n = 1 : \dots & & & \end{cases}$$

D'où on a tous les \widehat{V}_k , $k \geq 1$. Or les \widehat{V}_k , pour $k \leq 0$, sont nuls. On connaît donc tous les \widehat{V}_k .

— **Étape 2** : calcul des \widehat{X}_k , $k \leq 0$.

$$\begin{cases} \text{de la 1}^{\text{re}} \text{ équation : } & \widehat{X}_0 + \sum a_k X_k = \widehat{V}_0 + \sum b_l \widehat{V}_l & \Rightarrow & \widehat{X}_0 \\ \text{de la 2}^{\text{e}} \text{ équation : } & \widehat{X}_{-1} + \dots & \Rightarrow & \widehat{X}_{-1} \\ & \vdots & & \end{cases}$$

Pour $j < -q$,

$$\widehat{X}_j + \sum a_k \widehat{X}_{j+k} = 0$$

et donc $\widehat{X}_n = 0$ pour $n \leq s$, avec s grand.

— **Étape 3** : calcul des \widehat{U}_k , $k \leq N$.

$$\widehat{U}_n = \sum_0^\infty d_k \widehat{X}_k \quad \Rightarrow \quad \widehat{U}_n = 0 \quad \text{pour } n \leq s$$

et l'équation initiale

$$\sum_0^p a_k \widehat{X}_{n-k} = \sum_0^q b_l \widehat{U}_{n-l}$$

donne \widehat{U}_k pour $k \leq N$. Ensuite, on recalcule les \widehat{V}_k en fonction des $\widehat{U}_k \dots$

— **Étape 4** : calcul des \widehat{X}_k , $k > N$.

$$\widehat{X}_n + \sum_0^p a_k \widehat{X}_{n-k} = 0 \quad \text{dès que } n > N + p$$

Les \widehat{X}_k décroissent exponentiellement vite vers 0.

$$\widehat{X}_k = 0 \quad \text{pour } k > s', s' \text{ grand}$$

— **Étape 5** : on obtient de nouvelles valeurs pour les \widehat{V}_{N-p+j} , soient \widetilde{V}_{N-p+j} (correspondant à un nouveau vecteur $\Phi(\alpha)$). De la fonction

$$\begin{aligned} \mathbb{R}^d &\longrightarrow \mathbb{R}^d \\ \alpha &\longmapsto \Phi(\alpha) \end{aligned} \quad ,$$

il reste à déterminer le point fixe.

Proposition 2.7 — *Si $A_P X = A_Q U$ avec Q sublime, et si N est suffisamment grand, la suite $\Phi^{(r)}(\alpha)$ converge exponentiellement vite vers une limite α_∞ . De plus, les $\widehat{U}_k(r)$ calculés au cycle r convergent, pour chaque k , vers \widehat{U}_k . Par suite,*

$$F^{(r)}(a,b,\chi_N) \rightarrow F(a,b,\chi_N) .$$

2.4 Processus ARIMA

Il s'agit d'étudier la tendance, les périodicités, la non réversibilité et la variabilité non constante d'un processus. Pour parer à la variabilité non constante, on peut transformer la série au moyen d'une fonction déterministe. Concernant la tendance et les périodicités, on écrit

$$X_n = f(n) + V_n$$

avec V_n processus stationnaire, ou

$$X_n = p(n)$$

avec p périodique : $p(n) = p(n + T)$. Dans ce cas,

$$\frac{1}{T}(X_n + X_{n-1} + \dots + X_{n-T+1}) = \text{cte} + \frac{1}{T}(V_n + V_{n-1} + \dots + V_{n-T+1})$$

On note B l'opérateur de retard. Soit f un polynôme de degré $d - 1$:

$$(I - B)^d f = 0 .$$

Si $X_n = f(n) + V_n$, alors

$$(I - B)^d X_n = (I - B)^d V_n .$$

Exemple — $f(n) = \cos(2n\pi)/T$; alors

$$(I - B^T)f(n) = 0 .$$

On va utiliser $A_R X$, où

$$R(z) = \prod (z - z_i)^{s_i} ,$$

avec $|z_i| = 1$.

Si

$$R(z) = a_0 + a_1z + \cdots + a_dz^d,$$

alors

$$A_RX = a_0X_n + a_1X_{n-1} + \cdots + a_dX_{n-d}.$$

Définition 2.4 — Un processus du second ordre X est un **SARIMA** s'il existe un polynôme R n'ayant que des racines de module 1 tel que A_RX soit un ARMA.

Définition 2.5 — Un processus du second ordre X est un **ARIMA**(p, d, q) si $(I - B)^d X$ est un ARMA(p, q).

X SARIMA $\Leftrightarrow A_RX = Y$ ARMA(p, q), où

$$\begin{aligned} R(z) &= a_0 + a_1z + \cdots + a_dz^d, \\ A_Rf &= 0, \\ f &= \sum_{i=1}^d c_i f_i, \\ f_i(n) &= \left(\frac{1}{\tau_i}\right)^n, \end{aligned}$$

avec τ_i racine de R .

$$a_0X_n + \cdots + a_dX_{n-d} = Y_n.$$

La question qui se pose maintenant est la suivante : connaissant Y , comment déterminer X , modélisation de la série de départ ? On cherche une solution de $A_Rf = g$, avec $f, g : \mathbb{Z} \rightarrow \mathbb{R}$.

$$\phi * \psi(n) = \sum_{k=-\infty}^{+\infty} \phi(k)\psi(n-k).$$

La convolution commute les translations et la dérivation :

$$A_R(f * g) = (A_Rf) * g.$$

Si ψ est une solution de $A_Rf = g$, i.e. $A_R\psi = \delta_0$, alors

$$A_R(\psi * g) = (A_R\psi) * g = g.$$

Soient

$$\begin{aligned} \psi^+ &\text{ t.q. } \psi^+(n) = 0 \text{ pour } n < 0 \\ \psi^- &\text{ t.q. } \psi^-(n) = 0 \text{ pour } n \geq 0 \end{aligned}$$

Ce sont des solutions de $A_R\psi = \delta_0$.

$$\begin{cases} a_0\psi^+(n) + \dots + a_d\psi^+(n-d) = \delta_0(n) \\ a_0\psi^+(0) = 1 \quad \Rightarrow \psi^+(0) \\ a_0\psi^+(1) + a_1\psi^+(0) + 0 + \dots = 0 \quad \Rightarrow \psi^+(1) \\ \vdots \end{cases}$$

$$\begin{cases} \psi^-(n) = 0 \text{ pour } n > -d \\ a_0\psi^-(n) + \dots + a_d\psi^-(n-d) = \delta_0(n) \\ a_0\psi^-(-1) + a_1\psi^+(-1) + \dots + a_d\psi^-(-1) = \delta_0(-1) = 0 \\ \vdots \end{cases}$$

Notons $g^+(n) = g(n)\mathbf{1}_{(n \geq 0)}$ et $g^-(n) = g(n)\mathbf{1}_{(n < 0)}$.

$$\begin{aligned} A_R(\psi^+ * g^+) &= g^+ \\ A_R(\psi^- * g^-) &= g^- \end{aligned}$$

$$A_R(\psi^+ * g^+ + \psi^- * g^-) = g.$$

Proposition 2.8 — Si Y est un processus du second ordre, le processus \tilde{X} défini par

$$\tilde{X} = \psi^+ * Y^+ + \psi^- * Y^-$$

est du second ordre et est une solution de l'équation

$$A_RX = Y.$$

De plus, toutes les solutions de cette équation sont de la forme

$$\tilde{X}_n + c_1 f_1(n) + \dots + c_d f_d(n),$$

où les f_i forment une base de l'espace vectoriel des solutions de $A_R f = 0$ et les c_i sont des v.a. de carré intégrable.

Si $Y = A_Q W$, avec W BB, on a trouvé les solutions de

$$A_RX = A_Q W,$$

i.e. l'équation ARMA générale. Si R a toutes ses racines de module 1, alors

$$\frac{1}{R(z)} = \sum_0^\infty c_k z^k.$$

Soit $\psi^+(k) = c_k$: $(\sum a_k z^k)(\sum c_k z^k) = 1$. On fait l'hypothèse suivante :

(H) : les v.a. c_i sont orthogonales à H^Y .

Proposition 2.9 — Pour $n \geq 0$,

$$H_n^X = H_n^Y \oplus F ,$$

où

$$\begin{aligned} F &= ev(c_1, \dots, c_d) \\ &= ev(X_{-1}, \dots, X_{-d}) . \end{aligned}$$

2.5 Modèles multiplicatifs

On a vu que si le processus se met sous la forme $f(n) + U_n$, avec f de périodicité T , alors on calcule $(I - B^T)$ — et éventuellement $(I - B^T)^d$ — pour supprimer la périodicité. Soit une suite expérimentale x_1, \dots de période T . On la découpe en T séries :

$$\begin{aligned} &x_1, x_{1+T}, x_{1+2T}, \dots \\ &x_2, x_{2+T}, x_{2+2T}, \dots \\ &\vdots \\ &x_{T-1}, x_{2T-1}, x_{3T-1}, \dots \\ &x_T, x_{2T}, x_{3T}, \dots \end{aligned}$$

On pose $X_n^s = X_{s+nT}$. On fait l'hypothèse que la structure probabiliste ne dépend pas de s . Chacun de ces processus X^s est un ARIMA(p, d', q'). Ainsi,

$$\begin{aligned} \exists R, P, Q \text{ t.q. } R &= (I - B)^{d'} , \\ \forall s, A_{RP} X^s &= A_Q U^s . \end{aligned}$$

Attention — $U_n^s = U_{s+nT}$ n'est pas un bruit blanc.

Notation — Si P est un polynôme, on note

$$\tilde{P}(z) = P(z^T) .$$

On a alors

$$A_{\tilde{R}\tilde{P}} X = A_{\tilde{Q}} U ,$$

i.e. on a mis la saisonnalité dans les polynômes. U est un ARIMA(p, d, q), c.-à-d. qu'il existe ρ, π, χ tels que

$$A_{\rho\pi} U = A_\chi W ,$$

où W BB et $\rho(z) = (1 - z)^d$.

$$A_{\rho\pi\tilde{R}\tilde{P}} X = A_{\tilde{Q}\chi} W$$

On pose

$$Z = A_{\rho\tilde{R}}X.$$

Alors

$$A_{\rho\pi}Z = A_{\tilde{Q}\chi}W.$$

À ce processus ARMA, il correspond une unique solution stationnaire si les polynômes n'ont pas de racine de module 1. Z est stationnaire — c'est lui qu'on modélise.

Définition 2.6 — Une **SARIMA** $(p,d,q)(p',d',q')_T$ est un processus X tel que

$$(I - B)^d (I - B^T)^{d'} X$$

soit un ARMA $(p + p' + T, q + q' + T)$.

Exemple — Soit $T = 12$, $d = d' = 1$, $p = p' = q = q' = 1$. Nous sommes donc en présence de 5 coefficients (4 par les polynômes et un pour la variance du BB). C'est donc un ARMA(13,13).

Étape 1 : on cherche d et d' tels que $(I - B)^d (I - B^T)^{d'}$ soit « stationnaire ». On passe donc par $(I - B^T)$ pour obtenir $x_{13} - x_1$, $x_{14} - x_2$, ... Si la variance décroît rapidement vers 0, alors il s'agit d'un ARMA. Sinon, on passe par $(I - B)(I - B^T)$ pour obtenir $(x_{14} - x_2) - (x_{13} - x_1) - \dots$

Étape 2 : on regarde les $\hat{\gamma}$:

$$\hat{\gamma}(kT) \xrightarrow{\text{modélisation}} \text{ARMA}(P,Q).$$

Étape 3 : on regarde les $\hat{\gamma}$:

$$\hat{\gamma}(1), \hat{\gamma}(2), \dots, \hat{\gamma}(T-1) \xrightarrow{\text{modélisation}} \text{ARMA}(\pi,\chi)$$

2.6 Envoi

2.6.1 Critères de choix

Identification de (p,q) : quel est le meilleur ? est-ce que le meilleur est bon ?

2.6.2 Tests d'ajustement

1. Exemple d'un principe de critère de choix — Soit

$$\mathcal{E} = \mathbb{E}\left[\left(Y_{n+1} - (\hat{a}_1 Y_n + \dots + \hat{a}_p Y_{n-p+1})\right)^2\right],$$

$$Y_n - a_1 Y_{n-1} - \dots - a_p Y_{n-p} = U_n$$

(les a_i sont inconnus).

Les \hat{a}_i sont les coefficients estimés. D'où \mathcal{E} est l'erreur commise quand on prend les coefficients estimés.

$$\begin{aligned} \mathcal{E} &= \mathbb{E} \left[(U_{n+1} - ((\hat{a}_1 - a_1)Y_n + \dots + (\hat{a}_p - a_p)Y_{n-p+1}))^2 \right] \\ &= \sigma^2 + {}^t(\hat{a} - a) \Gamma_p (\hat{a} - a), \end{aligned}$$

où $\Gamma_p = \text{cov}(X_n, \dots, X_{n-p+1})$.

\hat{a} a été estimé à partir d'une série expérimentale x_1, \dots, x_N . Une bonne modélisation est une modélisation pour laquelle \mathcal{E} est petite. \mathcal{E} est une v.a. positive. Or

$$\sqrt{N}(\hat{a} - a) \longrightarrow \mathcal{N}(0, \sigma^2 \Gamma_p^{-1}).$$

D'où

$$\begin{aligned} \sqrt{N}(\hat{a} - a) {}^t(\sigma^2 \Gamma_p^{-1})^{-1} \sqrt{N}(\hat{a} - a) &\rightsquigarrow \chi^2(p) \\ \Rightarrow \mathcal{E} &\rightsquigarrow \sigma^2 + \frac{\sigma^2}{N} \chi^2(p) \\ \Rightarrow \mathbb{E}(\mathcal{E}) &= \sigma^2 \left(1 + \frac{p}{N}\right). \end{aligned}$$

Mais σ^2 est inconnu (il est lui aussi estimé : $\sigma^2 = \hat{\sigma}_p^2$, et il dépend de p). Quand $p \nearrow$, $1 + \frac{p}{N} \nearrow$, mais $\sigma^2 \searrow$. Il s'agit de chercher le p pour lequel $\mathbb{E}(\mathcal{E})$ est minimum. Pour un ARMA, on cherche à maximiser la vraisemblance. C'est le **critère d'Akaike**.

Exemple d'un principe de critère d'ajustement — Soit

$$A_P X = A_Q U, \quad \sigma_U^2$$

Soit x_1, \dots, x_N la réalisation de ce processus.

$$X_n + a_1 X_{n-1} + \dots + a_p X_{n-p} = U_n + b_1 U_{n-1} + \dots + b_q U_{n-q}$$

Alors

$$(w_n)_{n=p+1, \dots, N} = (x_n + a_1 x_{n-1} + \dots + a_p x_{n-p} - (U_n + b_1 U_{n-1} + \dots + b_q U_{n-q}))_{n=p+1, \dots, N}$$

est la réalisation d'un bruit blanc. Il existe diverses méthodes pour montrer que w_n est un BB.

Proposition 2.10 (Test du porte-manteau) — $\hat{\rho}_w(1), \hat{\rho}_w(2), \dots$ doivent être petits. La quantité

$$\sum_{k=1}^N \hat{\rho}_w^2(k)$$

suit un $\chi^2(N)$.

3

Modèles autorégressifs non linéaires

3.1 Rappels sur les modèles autorégressifs linéaires

3.1.1 Cadre univarié

Soit l'AR(1)

$$(\star) \quad X_n = aX_{n-1} + \epsilon_n$$

avec $a \in \mathbb{R}$, ϵ_n i.i.d. centrées et de variance $\sigma^2 \neq 0$. On cherche une solution telle que pour tout n , ϵ_n soit l'innovation, *i.e.* soit indépendant de $\sigma(X_p, p \leq n-1)$.

Notation — On note

$$\ln^+ x = \sup(0, \ln x).$$

Proposition 3.1 — *On suppose que $\epsilon_n \in \mathcal{L}^2$. Le modèle (\star) admet une solution stationnaire stricte dans \mathcal{L}^2 ssi $|a| < 1$.*

Proposition 3.2 — *Si $|a| < 1$, la solution stationnaire stricte du modèle est unique (et appartient à \mathcal{L}^2).*

Proposition 3.3 — *Si $|a| < 1$ et si $\mathbb{E}[\ln^+ |\epsilon_1|] < \infty$, alors le modèle admet une unique solution stationnaire stricte.*

Proposition 3.4 — *Soient V_n des v.a. i.i.d. positives.*

$$\begin{aligned} \mathbb{E}[\ln^+ V_1] < \infty &\Rightarrow \overline{\lim} V_n^{\frac{1}{n}} = 1, \\ \mathbb{E}[\ln^+ V_1] = \infty &\Rightarrow \overline{\lim} V_n^{\frac{1}{n}} = \infty. \end{aligned}$$

Lemme Soient V_n des v.a. i.i.d. positives.

$$\begin{aligned} \mathbb{P}(V_n = 0) < 1 &\Rightarrow \overline{\lim} V_n^{\frac{1}{n}} \geq 1 \text{ p.s. ,} \\ \mathbb{E} [\ln^+ V_1] < \infty &\Rightarrow \overline{\lim} V_n^{\frac{1}{n}} \leq 1 . \end{aligned}$$

Proposition 3.5 — Si $(a_n)_{n \geq 0}$ est une suite à valeurs dans \mathbb{R} et sous-additive (i.e. $a_{n+m} \leq a_n + a_m$), alors

$$\frac{a_n}{n} \xrightarrow[n \rightarrow \infty]{} \inf_n \frac{a_n}{n} .$$

Rappels —

$$\begin{aligned} \mathbb{E} [|X| < \infty] &\Leftrightarrow \sum_{n \geq 0} \mathbb{P}(|X| > n) < \infty , \\ \mathbb{E} [|X|] &= \int_0^\infty \mathbb{P}(|X| > t) dt . \end{aligned}$$

3.1.2 Cadre multivarié

Soit l'AR(1)

$$(\star\star) \quad X_n = AX_{n-1} + \epsilon_n ,$$

avec A matrice $d \times d$ et U_n une suite de vecteurs i.i.d. centrés de \mathcal{L}^2 . On cherche une solution telle que pour tout n , U_n soit l'innovation, i.e. soit indépendant de $\sigma(X_p, p \leq n-1)$.

Proposition 3.6 — Nous avons les résultats suivants :

1. si $(\star\star)$ admet une solution stationnaire stricte $(X_n)_{n \in \mathbb{Z}}$ qui est dans \mathcal{L}^2 et si la matrice de covariance K de $(X_n)_{n \in \mathbb{Z}}$ est inversible, alors les valeurs propres de A sont de modules inférieurs ou égaux à 1 ;
2. si la matrice de covariance Σ des U_n est inversible, alors les valeurs propres de A sont de modules strictement inférieurs à 1 ;
3. si les valeurs propres de A sont de modules 1, alors $(\star\star)$ admet une solution stationnaire stricte $(\bar{X}_n)_n$ dans \mathcal{L}^2 .

Rappel —

$$\begin{aligned} K &= \sum_{p \geq 0} A^p \Sigma (A^p)^t \\ &= \Sigma + \sum_{p \geq 1} A^p \Sigma (A^p)^t , \end{aligned}$$

et par suite,

$$K = \Sigma + AK A^t .$$

Définition 3.1 — Soit $\|\cdot\|$ une norme sur \mathbb{R}^d . On définit la **norme matricielle subordonnée** à $\|\cdot\|$ sur \mathbb{R}^d par

$$\|A\| = \sup_{\|v\|=1} \|Av\| .$$

Définition 3.2 — Le *rayon spectral* d'une matrice B est

$$\rho(B) = \sup \{ |\lambda_i| ; \lambda_i \text{ valeurs propres de } B \} .$$

Proposition 3.7 — Pour la norme

$$\|v\|_2 = \sqrt{\sum_{i=1}^d |v_i|^2}$$

sur \mathbb{R}^d , la norme matricielle subordonnée est

$$\|A\|_2 = \rho(AA^t)^{\frac{1}{2}} .$$

Corollaire 3.1 — Si A est symétrique ou diagonale, $\|A\|_2 = \rho(A)$.

Corollaire 3.2 — Nous avons les résultats suivants :

a)

$$\forall \beta > \rho(B), \exists \alpha, \forall n \in \mathbb{N}, \|B^n\| \leq \alpha \beta^n ;$$

b)

$$\rho(B) < 1 \Rightarrow B^n \rightarrow 0 ;$$

c)

$$\rho(B) < 1 \Leftrightarrow \exists n \text{ t.q. } \|B^n\| < 1 ;$$

d)

$$\left(\frac{\|B^n\|}{\beta^n} \right)^{\frac{1}{n}} \longrightarrow \frac{\rho(B)}{\beta} < 1 .$$

Propriété 3.1 —

$$\lim_n \|A^n\|^{\frac{1}{n}} = \inf_n \|A^n\|^{\frac{1}{n}} .$$

3.1.3 Retour au cadre univarié

Soit un AR(p) univarié :

$$(\star) \quad X_n = AX_{n-1} + U_n ,$$

avec

$$A = \begin{pmatrix} a_1 & \dots & \dots & \dots & a_p \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix} \quad \text{et} \quad U_n = \begin{pmatrix} \epsilon_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

Proposition 3.8 — Si (\star) admet une solution stationnaire stricte dans \mathcal{L}^2 et si $\mathbb{V}(\epsilon_n) \neq 0$, alors $\rho(A) < 1$.

Définition 3.3 — Le modèle (\star) est dit **commandable** si la matrice

$$\sum_{p \geq 0} A^p \Sigma (A^p)^t$$

existe et est inversible.

Proposition 3.9 — Soient A et Σ des matrices $d \times d$.

$$\sum_{p \geq 0} A^p \Sigma (A^p)^t \text{ inversible} \iff \sum_{p=0}^{d-1} A^p \Sigma (A^p)^t \text{ inversible}.$$

Proposition 3.10 — Soit le modèle (\star) . Alors il existe une solution stationnaire \mathcal{L}^2 ssi $\rho(A) < 1$.

3.2 Modèles autorégressifs non linéaires lipschitziens

Soit le modèle

$$(\star) \quad X_n = F(X_{n-1}, \epsilon_n) = F_{\epsilon_n}(X_{n-1}),$$

avec X_n à valeurs dans \mathbb{R}^d et ϵ_n i.i.d. On cherche une solution telle que pour tout n , ϵ_n soit l'innovation, i.e. soit indépendant de $\sigma(X_p, p \leq n-1)$.

Définition 3.4 — La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est **lipschitzienne** de coefficient de lipschitz

$$c_f = \sup \frac{\|f(x) - f(y)\|_p}{\|x - y\|_q}$$

si $c_f < \infty$.

Théorème 3.1 — Soit le modèle (\star) avec les ϵ_i i.i.d. et pour tout n , ϵ_n indépendant de $\sigma(X_p, p \leq n-1)$. X_n est à valeurs dans \mathbb{R}^k muni de la norme $\|\cdot\|$. On suppose que $\mathbb{E}[\ln^+ c_{F_{\epsilon_1}}] < \infty$ et qu' $\exists x \in \mathbb{R}^k$ tel que $\mathbb{E}[\ln^+ \|F_{\epsilon_1}(x) - x\|] < \infty$. Alors :

- 1) si $\mathbb{E}[\ln c_{F_{\epsilon_1}}] < 0$, il existe une unique solution \bar{X}_n du modèle qui est stationnairement stricte ;
- 2) s'il existe $k \in \mathbb{N}^*$ t.q. $\mathbb{E}[c_{F_{\epsilon_1}}^k] < 1$ et si $\mathbb{E}[\|F_{\epsilon_1}(x) - x\|^k] < \infty$, alors la solution \bar{X}_n stationnaire stricte a un moment d'ordre k .

Nota —

$$\begin{aligned} X_n &= F_{\epsilon_n}(X_{n-1}) \\ &= F_{\epsilon_n} \circ F_{\epsilon_{n-1}}(X_{n-2}) \\ &= F_{\epsilon_n} \circ \dots \circ F_{\epsilon_{n-p+1}}(X_{n-p}) . \end{aligned}$$

D'où

$$\bar{X}_n = \lim_{p \rightarrow \infty} F_{\epsilon_n} \circ \dots \circ F_{\epsilon_{n-p+1}}(0) .$$

Théorème 3.2 — Soit le modèle (\star) avec les ϵ_i i.i.d. et pour tout n , ϵ_n indépendant de $\sigma(X_p, p \leq n-1)$. X_n est à valeurs dans \mathbb{R}^k muni de la norme $\|\cdot\|$. On suppose que $\mathbb{E}[\ln^+ c_{F_{\epsilon_1}}] < \infty$ et qu' $\exists x \in \mathbb{R}^k$ tel que $\mathbb{E}[\ln^+ \|F_{\epsilon_1}(x) - x\|] < \infty$. Alors :

- 1) s'il existe p t.q. $\mathbb{E}[\ln c_{F_{\epsilon_1} \circ \dots \circ F_{\epsilon_p}}] < 0$, alors il existe une unique solution stationnaire stricte au modèle;
- 2) s'il existe $k, p \in \mathbb{N}$ t.q. $\mathbb{E}[c_{F_{\epsilon_1} \circ \dots \circ F_{\epsilon_p}}^k] < 1$ et t.q. $\mathbb{E}[\|F_{\epsilon_1} \circ \dots \circ F_{\epsilon_p}(x) - x\|^k] < \infty$, alors il existe une unique solution \bar{X}_n stationnaire stricte, et cette solution a des moments d'ordre k .

Notation —

$$X_n^x = \begin{cases} X_n = F(X_{n-1}, \epsilon_n) \\ X_0 = x \end{cases}$$

Théorème 3.3 — S'il existe $k, p \in \mathbb{N}^*$ t.q. $\mathbb{E}[\|F_{\epsilon_1}(x) - x\|^k] < \infty$ et $s \exists \beta < 1$ t.q. $\mathbb{E}[\|X_p^x - X_p^y\|^k] \leq \beta \|x - y\|^k, \forall x, \forall y$, alors il existe une solution stationnaire stricte, et cette solution a des moments d'ordre k .

Lemme 3.1 — Sous les hypothèses du théorème précédent,

$$\mathbb{E}[\|F_{\epsilon_n} \circ \dots \circ F_{\epsilon_{n-r+1}}\|^k] \leq \beta^r \|x - y\|^k ,$$

ce qui équivaut à

$$\mathbb{E}[\|X_r^x - X_r^y\|^k] \leq \beta^r \|x - y\|^k .$$

3.2.1 Modèles hétéroscédastiques

Soit

$$X_n = f(X_{n-1}) + g(X_{n-1})\epsilon_n .$$

$$F_{\epsilon_1}(x) - F_{\epsilon_1}(y) = f(x) + g(x)\epsilon_1 - f(y) - g(y)\epsilon_1, \text{ d'où } c_{F_{\epsilon_1}} \leq c_f + c_g|\epsilon_1| .$$

Nous avons que :

- si $\mathbb{E}[\ln c_{F_{\epsilon_1}}] \leq \mathbb{E}[\ln(c_f + c_g|\epsilon_1|)] < 0$, alors il existe une solution strictement stationnaire;

- si $\mathbb{E}[c_f + c_g | \epsilon_1] < 1$ et si $c_f + c_g \mathbb{E}(|\epsilon_1|) < 1$, alors il existe une solution strictement stationnaire ayant un moment d'ordre 1 ;
- si $c_f^2 + c_g^2 \mathbb{E}(\epsilon_1^2) + 2c_f c_g \mathbb{E}(|\epsilon_1|) < 1$, alors il existe une solution strictement stationnaire ayant un moment d'ordre 2 ;
- si $c_f^2 + c_g^2 \mathbb{E}(\epsilon_1^2) < 1$, alors il existe une solution strictement stationnaire ayant un moment d'ordre 2 ;
- si $\mathbb{E}(\epsilon_1) = 0$, $\mathbb{E}(\epsilon^2) = 1$ et si $c_f^2 + c_g^2 < 1$, alors il existe une solution strictement stationnaire ayant un moment d'ordre 2.

3.2.2 Modèle autorégressif non linéaire à coefficients aléatoires

Soit

$$X_n = A(\epsilon_n)X_{n-1} + B(\epsilon_n) = F(X_{n-1}, \epsilon_n).$$

Proposition 3.11 — *On suppose que $\mathbb{E}[\ln \|A(\epsilon_1)\|] < \infty$. Alors*

$$\frac{1}{p} \mathbb{E}[\ln \|A(\epsilon_1) \times \dots \times A(\epsilon_p)\|] \longrightarrow \gamma = \inf_p \frac{1}{p} \mathbb{E}[\ln \|A(\epsilon_1) \times \dots \times A(\epsilon_p)\|].$$

Si $\gamma < 0$, $\exists p$ t.q. $\mathbb{E}[\ln \|A(\epsilon_1) \times \dots \times A(\epsilon_p)\|] < 0$, et alors le modèle a une solution stationnaire stricte si $\mathbb{E}[\ln^+ \|B(\epsilon_1)\|] < \infty$.

Définition 3.5 — *γ est appelé le plus grand exposant de Lyapounov du produit des matrices aléatoires.*

Proposition 3.12 — *Nous avons*

$$\frac{1}{p} \ln \|A(\epsilon_1) \times \dots \times A(\epsilon_p)\| \longrightarrow \gamma \quad p.s. ,$$

i.e.

$$\|A(\epsilon_1) \times \dots \times A(\epsilon_p)\|^{\frac{1}{p}} \longrightarrow e^\gamma \quad p.s. .$$

3.3 Ergodicité

Définition 3.6 — *On a un processus $(X_n)_n$ indexé par \mathbb{N} ou \mathbb{Z} , $X_n : (\Omega, \mathcal{F}, \mathbb{P})$. On lui associe le processus canonique (\hat{X}_n) indexé par \mathbb{N} ou \mathbb{Z} et défini sur $(E^{\mathbb{N}}$ ou \mathbb{Z} , $\mathcal{B}(E^{\mathbb{N}}$ ou \mathbb{Z}). On considère*

$$\begin{aligned} \phi : \Omega &\longrightarrow E^{\mathbb{N}} \\ \omega &\longmapsto \phi(\omega) = (X_0(\omega), \dots, X_n(\omega)) \\ &\quad \text{(trajectoires du processus)}. \end{aligned}$$

On note (\widehat{X}_n) la n^e application coordonnée de $E^{\mathbb{N}}$.

(\widehat{X}_n) est défini sur $(E^{\mathbb{N}}, \mathcal{B}(E^{\mathbb{N}}), \widehat{\mathbb{P}})$ où $\widehat{\mathbb{P}}$ est l'image de \mathbb{P} par ϕ . C'est le processus canonique associé à $(X_n)_n$.

On peut définir le **shift** θ sur $E^{\mathbb{N}}$:

$$\begin{aligned}\theta(x_0, x_1, \dots, x_n, \dots) &= (x_1, x_2, \dots, x_{n+1}, \dots), \\ \widehat{X}_n \circ \theta &= \widehat{X}_{n+1}, \\ \widehat{X}_n \circ \theta^p &= \widehat{X}_{n+p}.\end{aligned}$$

Proposition 3.13 — (\widehat{X}_n) est stationnaire strict ssi

$$\theta\widehat{\mathbb{P}} = \widehat{\mathbb{P}}.$$

Définition 3.7 — On appelle **tribu des invariants** associée à (\widehat{X}_n) l'ensemble

$$\mathfrak{I} = \{A \in \mathcal{B}(E^{\mathbb{N}}), \theta^{-1}(A) = A\}.$$

On appelle tribu des invariants associée à (X_n)

$$\bar{\mathfrak{I}} = \phi^{-1}(\mathfrak{I}).$$

Définition 3.8 — Un processus (X_n) est dit **ergodique** si sa tribu des invariants associée est p.s. grossière, i.e.

$$\forall A \in \bar{\mathfrak{I}}, \mathbb{P}(A) = 0 \text{ ou } 1.$$

Remarque — Nous avons

$$\widehat{X}_n \text{ ergodique} \Leftrightarrow X_n \text{ ergodique}.$$

Proposition 3.14 —

$$A \in \bar{\mathfrak{I}} \Leftrightarrow \exists B \in \mathcal{B}(E^{\mathbb{N}}), \forall n, A = \left\{ \omega \mid (X_n(\omega), X_{n+1}(\omega), \dots) \in B \right\}.$$

Définition 3.9 — La **tribu asymptotique**, pour un processus $(X_n)_n$, est

$$\mathcal{A}_{\infty} = \bigcap_p \sigma(X_n, n \geq p).$$

Corollaire 3.3 —

$$\bar{\mathfrak{I}} \subset \mathcal{A}_{\infty}.$$

Proposition 3.15 — Si $(\epsilon_n)_n$ est une suite de v.a. indépendantes et de même loi, alors $(\epsilon_n)_n$ est stationnaire stricte et ergodique.

Proposition 3.16 — Soit $(X_i)_i$ une suite stationnaire stricte et ergodique. Soit $\phi : E^{\mathbb{N}} \rightarrow \tilde{E}$. On pose, $\forall n \in \mathbb{E}$, $\tilde{X}_i = \phi(X_i, X_{i+1}, \dots, X_{i+n}, \dots)$. Alors \tilde{X}_i est stationnaire stricte ergodique.

Proposition 3.17 — Si $(\epsilon_n)_{n \in \mathbb{Z}}$ est une suite de v.a. indépendantes et de même loi, alors $(\epsilon_n)_{n \in \mathbb{Z}}$ est stationnaire stricte et ergodique.

Lemme 3.2 — La tribu des invariants est incluse p.s. dans la tribu asymptotique

$$\bigcap_p \sigma(\epsilon_p, p \geq n).$$

Conséquences — Elles sont au moins au nombre de 3 :

- 1) si les $(\epsilon_n)_n$ sont i.i.d., alors $(\epsilon_n)_n$ est stationnaire stricte ergodique ;
- 2) si $X_n = \psi(\epsilon_n, \epsilon_{n-1}, \dots, \epsilon_{n-k}, \dots)$, alors $(X_n)_n$ est stationnaire stricte ergodique ;
- 3) si $(a_i)_i$ est t.q. $\sum |a_i|^2 < \infty$, alors $X_n = \sum a_i \epsilon_{n-i}$ stationnaire stricte ergodique.

Théorème 3.4 (Birkoff) — Soit $(X_i)_{i \in \mathbb{Z}}$ un processus stationnaire strict. Alors

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} X_i = \mathbb{E}(X_0 | \mathfrak{I}) \quad p.s. .$$

Si X_0 est intégrable, et si de plus le processus est ergodique, alors

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} X_i = \mathbb{E}(X_0) \quad p.s. .$$

Lemme 3.3 (Ergodicité maximale) — Soit $(X_n)_n$ un processus stationnaire strict tel que $\mathbb{E}(|X_0|) < \infty$. Soit $S_n = X_0 + X_1 + \dots + X_{n-1}$. On pose $M_n = \max(0, S_1, \dots, S_n) \geq 0$. Alors

$$\int_{\{M_n > 0\}} X_0 d\mathbb{P} \geq 0.$$

Théorème 3.5 (Ergodicité sous-additive) —

$$U_{n+m} \leq U_m + U_n \circ \theta^n \quad \Rightarrow \quad \frac{U_n}{n} \text{ converge p.s. .}$$

Proposition 3.18 — La convergence du théorème de Birkoff a aussi lieu dans \mathcal{L}^1 .

3.4 Chaînes de Markov et stabilité

Soit $(X_n)_n$ une chaîne de Markov homogène de probabilité de transition π :

$$\begin{aligned} \forall B \in \mathcal{B}(E), \quad \mathbb{P}(X_{n+1} \in B \mid \mathcal{F}_n) &= \mathbb{P}(X_{n+1} \in B \mid X_n) \\ &= \pi(X_n, B). \end{aligned}$$

Soit ν la loi initiale de cette chaîne.

Proposition 3.19 — *La chaîne de Markov est stationnaire stricte ssi*

$$\nu\pi = \nu.$$

Proposition 3.20 — *Étant données une proba de transition π et une loi initiale ν , il existe sur $(E^{\mathbb{N}}, \mathcal{B}(E^{\mathbb{N}}))$ où X_n est la n^e application coordonnée, une unique loi de probabilité \mathbb{P}_ν telle que $\forall A_i \in \mathcal{B}(E^{\mathbb{N}})$,*

$$\mathbb{P}_\nu(X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n) = \int_{A_0} \nu(dx_0) \int_{A_1} \pi(x_0, dx_1) \dots \int_{A_n} \pi(x_{n-1}, dx_n).$$

Notation — Nous notons

$$\nu = \delta_x \quad \implies \quad \mathbb{P}_\nu = \mathbb{P}_x.$$

Définition 3.10 — *La chaîne de Markov (X_n) est dite **stable** s'il existe une proba μ telle que*

$$\forall f \in C_b, \forall x \in E, \quad \frac{1}{n} \sum_{i=1}^n f(X_i) \longrightarrow \int d\mu \quad \mathbb{P}_x - p.s..$$

Définition 3.11 — *La chaîne de Markov (X_n) de proba de transition π est dite **fellerienne** si*

$$\forall f \in C_b, \quad \pi f \in C_b.$$

Proposition 3.21 — *Si la chaîne de Markov de proba de transition π est fellerienne et stable, la loi limite μ vérifie*

$$\mu\pi = \mu.$$

Proposition 3.22 — *Soit le modèle $X_n = F(X_{n-1}, \epsilon_n)$. Si ce modèle admet une solution stationnaire stricte et ergodique \bar{X}_n et si $\forall x, X_n^x - \bar{X}_n \rightarrow 0$ p.s., alors $\forall f$ uniformément continue, $\forall x$,*

$$\frac{1}{n} \sum_{i=1}^n f(X_i^x) \xrightarrow{p.s.} \int f(x) d\nu(x),$$

où ν est la loi de \bar{X}_1 .

On a, en plus du résultat de la proposition précédente, que $\forall f$ continue bornée, $\forall x$,

$$P^n f(x) \longrightarrow \int f d\mu \quad (n \rightarrow \infty).$$

ce qui équivaut à

$$\forall x, \quad P^n(x, \cdot) \xrightarrow{\text{étroit.}} \mu.$$

Corollaire 3.4 — Soit le modèle $X_n = F(X_{n-1}, \epsilon_n) = F_{\epsilon_n}(X_{n-1})$. S'il existe x tel que :

- $\mathbb{E} [\ln^+ \|F_{\epsilon_1}(x) - x\|] < \infty$;
- $\mathbb{E} [\ln^+ c_{F_{\epsilon_1}}] < \infty$;
- $\mathbb{E} [\ln c_{F_{\epsilon_1}}] < 0$.

Alors il existe une solution stationnaire stricte ergodique et la chaîne de Markov associée est stable.

Définition 3.12 — Une chaîne de Markov est dite **récurrente positive** s'il existe une proba μ telle que $\forall f$ bornée, $\forall x \in \mathbb{R}^d$,

$$\frac{1}{n} \sum_{i=1}^n f(X_i^x) \xrightarrow{p.s.} \int f(x) d\mu(x),$$

ce qui revient à dire que $\forall A$ borélien tel que $\mu(A) > 0$, partant de tout point x , la chaîne visite une infinité de fois A .

Proposition 3.23 — Si la chaîne est stable, alors $\forall O$ ouvert de mesure $\mu(O) > 0$, $\forall x \in \mathbb{R}^d$, la chaîne issue de x visite p.s. une infinité de fois l'ouvert O . On dit qu'il y a **récurrence dans les ouverts chargés par la proba invariante**.

Définition 3.13 — Une proba de transition P est dite **fortement fellerienne** si $\forall f$ bornée, Pf est continue et bornée.

Exemple — Soit

$$X_{n+1} = f(X_n) + \epsilon_{n+1}.$$

Si f est continue, P est fellerienne. Si f est continue et si ϵ_1 admet une densité par rapport à la mesure de Lebesgue, alors P est fortement fellerienne.

Proposition 3.24 — Si $(X_n)_n$ est stable et P fortement fellerienne, alors $(X_n)_n$ est récurrente positive.

Proposition 3.25 (Fonction de Lyapounov) — Soit $V : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction continue telle que $V(x) \rightarrow \infty$ quand $\|x\| \rightarrow \infty$. Si $\nu_n \rightarrow \nu$ étroitement et si $\nu_n(V) \rightarrow \nu(V)$, alors $\forall \Phi$ à valeurs réelles, continue et telle que $|\Phi| \leq \alpha V + \beta$,

$$\nu_n(\Phi) \longrightarrow \nu(\Phi).$$

Définition 3.14 — V est appelée *fonction de Lyapounov*.

Application 1 — Si on sait que

$$\frac{1}{n} \sum_{i=1}^n V(X_i^x) \longrightarrow \int V d\mu$$

et que V est continue et positive, alors $\forall \Phi$ continue telle que $|\Phi| \leq \alpha V + \beta$,

$$\frac{1}{n} \sum_{i=1}^n \Phi(X_i^x) \longrightarrow \int \Phi d\mu .$$

Application 2 — Si le modèle $X_{n+1} = F(X_n, \epsilon_{n+1})$ admet une solution stationnaire stricte $(\bar{X}_n)_n$ qui a un moment d'ordre 1 et qui est ergodique, et si $\bar{X}_n - X_n^x \rightarrow 0$ p.s., alors $\forall \Phi$ à valeurs réelles, continue et telle que $|\Phi| \leq \alpha V + \beta$,

$$\frac{1}{n} \sum_{i=1}^n \Phi(X_i^x) \xrightarrow{\text{p.s.}} \int \Phi d\nu .$$

Application 3 — Soit le modèle $X_n = aX_{n-1} + U_n$ avec $\rho(A) < 1$; si U_n a un moment d'ordre 2, alors il existe une solution stationnaire stricte ergodique ayant un moment d'ordre 2.

3.5 Modèles ARCH et GARCH

Un modèle AR classique s'écrit

$$X_n = a_1 X_{n-1} + \dots + a_p X_{n-p} + \epsilon_n ,$$

avec les ϵ_n bruit blanc gaussien. Ici, il s'agit de modéliser autrement l'erreur. Soit le modèle

$$(\star) \quad \begin{cases} \epsilon_n = \sqrt{h_{n-1}} \cdot \eta_n \\ \eta_n \text{ i.i.d., } \mathbb{E}(\eta_n) = 0, \mathbb{E}(\eta_n^2) = 1 \\ \eta_n \text{ indépendant de } \epsilon_{n-1} = \sigma(\epsilon_p, p \neq n-1) \\ h_{n-1} \text{ } \sigma(\epsilon_p, p \neq n-1) \text{ - mesurable} \end{cases}$$

Nous avons que

$$\begin{aligned} \mathbb{E}(\epsilon_n | \epsilon_{n-1}) &= \sqrt{h_{n-1}} \cdot \mathbb{E}(\eta_n) = 0 , \\ \mathbb{E}(\epsilon_n^2 | \epsilon_{n-1}) &= h_{n-1} \cdot \mathbb{E}(\eta_n^2) = h_{n-1} . \end{aligned}$$

Définition 3.15 — Le modèle est dit *hétéroscédastique* si $\mathbb{E}(\epsilon_n^2 | \epsilon_{n-1})$ n'est pas constant.

Proposition 3.26 — Si ϵ_n solution du modèle (\star) vérifie $\mathbb{E}(\epsilon_n^2) = \text{cste} < \infty$, alors ϵ_n est un bruit blanc de \mathcal{L}^2 (i.e. bruit blanc au sens faible).

Définition 3.16 — Le modèle (\star) est un **ARCH(q)** (autoregressive conditionally heteroshedastical) si

$$h_{n-1} = \gamma + \sum_{i=1}^q \alpha_i \epsilon_{n-i}^2,$$

avec $\gamma \geq 0, \alpha_i \geq 0 \quad \forall i$.

Définition 3.17 — Le modèle (\star) est un **GARCH(p,q)** (generalised autoregressive conditionally heteroshedastical) si

$$h_{n-1} = \gamma + \sum_{i=1}^q \alpha_i \epsilon_{n-i}^2 + \sum_{j=1}^p \beta_j h_{n-j-1},$$

avec $\gamma \geq 0, \alpha_i, \beta_i \geq 0 \quad \forall i$.

Proposition 3.27 — S'il existe une solution ϵ_n du modèle ARCH(q) telle que ϵ_n^2 soit stationnaire au sens large, alors ϵ_n^2 est un AR(q) vérifiant

$$\epsilon_n^2 = \gamma + \sum_{i=1}^q \alpha_i \epsilon_{n-i}^2 + U_n,$$

avec U_n bruit blanc faible.

S'il existe une solution ϵ_n du modèle GARCH(p,q) telle que ϵ_n^2 soit stationnaire au sens large, alors ϵ_n^2 est un ARMA(sup(p,q),q) vérifiant

$$\epsilon_n^2 = \gamma + \sum_{i=1}^{\text{sup}(p,q)} (\alpha_i + \beta_i) \epsilon_{n-i}^2 - \sum_{j=1}^p \beta_j U_{n-j} + U_n,$$

avec U_n bruit blanc faible.

Proposition 3.28 — Si le modèle (\star) admet une solution stationnaire faible, alors $\alpha + \beta < 1$.

Proposition 3.29 — Si le modèle (\star) admet une solution stationnaire stricte, alors $\mathbb{E}[\ln(\alpha + \beta \eta_1^2)] < 0$. Dans ce cas, la solution est ergodique. de plus, si la solution stationnaire stricte admet un moment d'ordre 2, alors $\alpha + \beta < 1$.

Proposition 3.30 — Si le modèle GARCH(p,q) admet une solution stationnaire faible, alors

$$\sum_{i=1}^{\text{sup}(p,q)} (\alpha_i + \beta_i) < 1.$$

Proposition 3.31 — Soit $(A_n)_{n \in \mathbb{Z}}$ une suite de matrices aléatoires formant un processus stationnaire strict. Alors

$$\mathbb{E} [\ln \|A_1\|] < \infty \iff \frac{\mathbb{E} [\ln \|A_1 \dots A_n\|]}{n} \longrightarrow \gamma ,$$

où γ est le plus grand exposant de Lyapounov. De plus,

$$\frac{\ln \|A_1 \dots A_n\|}{n} \xrightarrow{p.s.} \gamma ,$$

i.e.

$$\|A_1 \dots A_n\|^{\frac{1}{n}} \xrightarrow{p.s.} e^\gamma .$$

Proposition 3.32 — Si $\gamma < 0$, alors il existe une solution stationnaire stricte.

Proposition 3.33 — Le modèle GARCH(p, q) admet une solution stationnaire stricte dans \mathcal{L}^2 ssi

$$\sum_{i=1}^{\sup(p,q)} (\alpha_i + \beta_i) < 1 .$$

3.6 Modèles de diffusions limites des modèles GARCH

Soit le modèle de diffusion

$$(\star) \quad \begin{cases} dY_t = b(Y_t) dt + \sigma(Y_t) dW_t \\ Y_0 = y_0 \end{cases}$$

La **discrétisation d'Euler** est définie par

$$Y_{(k+1)h} - Y_{kh} = b(Y_{kh}) \times h + \underbrace{\sigma(Y_{kh})(W_{(k+1)h} - W_{kh})}_{\sqrt{h}Z_{k+1}^h}$$

$Z_k = Y_{kh}$ est une chaîne de Markov.

$$\begin{aligned} \frac{1}{h} \mathbb{E} [Y_{(k+1)h} - Y_{kh} \mid Y_{kh} = y] &= b(y) \\ \frac{1}{h} \mathbb{V} [Y_{(k+1)h} - Y_{kh} \mid Y_{kh} = y] &= \sigma^2(y) \end{aligned}$$

Théorème 3.6 (Stroock - Varadhan) — Soit $(Y_k^{(h)})_k$ une famille de chaînes de Markov indexées par h , à valeurs dans \mathbb{R}^d . $\bar{Y}_t^h = Y_k^h$ si $t \in [kh, (k+1)h[$, $t \in \mathbb{R}$.

C'est un processus cadlag. Supposons qu' $\exists a, b$, applications continues, avec $a(y)$ matrice d -dimensionnelle définie positive et $b(y)$ vecteur de \mathbb{R}^d . On suppose que

$$\begin{aligned} \sup_{|y| \leq r} \left| \frac{1}{h} \mathbb{E} [Y_1^{(h)} - Y_0^{(h)} \mid Y_0^{(h)} = y] - b(y) \right| &\longrightarrow 0 \quad (h \rightarrow 0) , \\ \sup_{|y| \leq r} \left| \frac{1}{h} \text{Cov} [Y_1^{(h)} - Y_0^{(h)} \mid Y_0^{(h)} = y] - a(y) \right| &\longrightarrow 0 \quad (h \rightarrow 0) , \\ \sup_{|y| \leq r} \left| \frac{1}{h^{1+\delta/2}} \mathbb{E} [|Y_1^{(h)} - Y_0^{(h)}|^{2+\delta} \mid Y_0^{(h)} = y] \right| &\text{reste borné quand } h \rightarrow 0 . \end{aligned}$$

S'il existe σ continue telle que $a(y) = \sigma(y)\sigma(y)^t$ et si l'EDS (\star) admet une solution unique, alors les lois fini-dimensionnelles des processus \bar{Y}_t^h convergent vers celle de l'EDS.

Conséquence — Discrétisation d'Euler $\xrightarrow{h \rightarrow 0}$ solution de l'EDS.

Deuxième partie

THÓRIE DE MARKOV

4

Introduction

Définition 4.1 — Un **processus de Markov** est un processus tel que, étant donné la valeur de X_t , la valeur de X_s pour $s > t$ ne dépend pas des valeurs prises avant t , soient $\{X_u, u < t\}$.
Ceci s'écrit

$$\mathbb{P}(X_t \in]a, b] \mid X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n) = \mathbb{P}(X_t \in]a, b] \mid X_{t_n} = x_n)$$

pour $t_1 < t_2 < \dots < t_n < t$.

Définition 4.2 — On appelle **fonction de probabilité de transition** la fonction

$$\mathbb{P}(x, s; t, A) = \mathbb{P}(X_t \in A \mid X_s = x)$$

pour $t > s$ et $A \subset \mathbb{R}$.

Définition 4.3 — Un processus de Markov ayant un espace d'états fini ou dénombrable est appelé **chaîne de Markov**. Un processus de Markov pour lequel toutes les réalisations $\{X_t, t \in [0, \infty[\}$ sont des fonctions continues est appelé **processus de diffusion**.

Définition 4.4 — Un processus est dit **stationnaire** si, pour tout $h > 0$,

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}) \stackrel{\mathcal{L}}{=} (X_{t_1}, X_{t_2}, \dots, X_{t_n}).$$

Définition 4.5 — Un processus est dit **stationnaire par covariance** si ses moments du second ordre sont finis et si

$$\text{Cov}(X_t, X_{t+h}) = \mathbb{E}(X_t X_{t+h}) - \mathbb{E}(X_t) \cdot \mathbb{E}(X_{t+h})$$

ne dépend que de h pour tout $t \in T$.

Proposition 4.1 — *Un processus stationnaire ayant ses moments d'ordre 2 finis est un processus stationnaire par covariance.*

Définition 4.6 — *Un processus a ses probabilités de transition stationnaires si $\mathbb{P}(x, s ; t, A)$ ne dépend que de $t - s$.*

Notation — On note

$$P_{ij}^{n, n+m} = \mathbb{P}(X_{n+m} = j \mid X_n = i)$$

et

$$\begin{aligned} P_{ij}^{n, n+1} &= P_{ij} \\ &= \mathbb{P}(X_{n+1} = j \mid X_n = i) . \end{aligned}$$

Définition 4.7 — $P_{ij}^{n, n+1}$ est appelée **probabilité de transition en un pas (one-step transition)**, et $P_{ij}^{n, n+m} = P_{ij}^m$ est appelée **probabilité de transition en m pas (m-step transition)**.

Proposition 4.2 — *Nous avons*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p_{i_0} \cdot P_{i_0, i_1} \dots P_{i_{n-2}, i_{n-1}} \cdot P_{i_{n-1}, i_n}$$

avec $p_{i_0} = \mathbb{P}(X_0 = i_0)$.

Remarque — Une chaîne de Markov est déterminée par sa matrice de probabilité de transition et la distribution de probabilité du processus à l'instant 0.

Définition 4.8 — Une **marche aléatoire uni-dimensionnelle** est une chaîne de Markov d'espace d'états l'ensemble (fini ou infini) $\{a, a + 1, \dots, b\}$ pour lequel, si le processus est en i à l'instant n , alors à l'instant $n + 1$ il ne peut être qu'en i , en $i - 1$ ou en $i + 1$. La matrice de transition est alors de la forme

$$\begin{pmatrix} r_0 & p_0 & 0 & 0 & \dots & \dots & 0 \\ q_1 & r_1 & p_1 & 0 & \dots & \dots & 0 \\ 0 & q_2 & r_2 & p_2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & r_b \end{pmatrix}$$

avec $p_i > 0$, $q_i > 0$, $r_i \geq 0$ et $q_i + r_i + p_i = 1$ pour tout $i \in \{1, 2, \dots\}$, $p_0 \geq 0$, $r_0 \geq 0$, $r_0 + p_0 = 1$ et enfin, si $X_n = i$, $i \geq 1$,

$$\begin{cases} \mathbb{P}(X_{n+1} = i + 1 \mid X_n = i) = p_i , \\ \mathbb{P}(X_{n+1} = i \mid X_n = i) = r_i , \\ \mathbb{P}(X_{n+1} = i - 1 \mid X_n = i) = q_i . \end{cases}$$

Proposition 4.3 — Si la matrice de probabilité de transition en un pas d'une chaîne de Markov est P , alors

$$P_{ij}^n = \sum_{k=0}^{\infty} P_{ik}^r \cdot P_{kj}^s$$

pour toute paire (r,s) d'entiers positifs vérifiant $r + s = n$, et avec la convention

$$P_{ij}^0 = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Définition 4.9 — Un état j est dit **accessible** à partir d'un état i s'il existe un entier $n > 0$ tel que $P_{ij}^n > 0$.

Définition 4.10 — Deux états i et j **communiquent** s'ils sont mutuellement accessibles. On note cette communicabilité $i \leftrightarrow j$.

Proposition 4.4 — Le critère de communicabilité est une relation d'équivalence :

- (i) $i \leftrightarrow i$ (réflexivité);
- (ii) $i \leftrightarrow j \Rightarrow j \leftrightarrow i$ (symétrie);
- (iii) $i \leftrightarrow j$ et $j \leftrightarrow k \Rightarrow i \leftrightarrow k$ (transitivité).

Il est par conséquent possible de partitionner l'ensemble des états en classes d'équivalence.

Par ailleurs, s'il était possible, partant d'une classe, d'entrer dans une autre classe avec une probabilité positive, alors il serait clairement impossible de retourner dans la classe initiale, à moins que ces deux classes n'en forment qu'une seule.

Définition 4.11 — Une chaîne de Markov est dite **irréductible** si la relation d'équivalence induit une seule classe, i.e. tous ses états communiquent entre eux.

Définition 4.12 — La **période** d'un état i , noté $d(i)$, est le plus grand commun diviseur (pgcd) de tous les entiers $n \geq 1$ pour lesquels $P_{ii}^n > 0$. Par convention, on pose $d(i) = 0$ si $P_{ii}^n = 0$ pour tout $n \geq 1$.

Remarques — Nous avons :

1. Si, pour une marche aléatoire, $r_i = 0$ quel que soit i , alors tous les états de cette marche aléatoire ont pour période 2.
2. Si, pour une marche aléatoire, il existe un état i_0 tel que $r_{i_0} > 0$, alors tous les états de cette marche aléatoire ont pour période 1.

Théorème 4.1 — $i \leftrightarrow j \Rightarrow d(i) = d(j)$.

Remarque — Ceci prouve que la périodicité est une propriété de classe.

Théorème 4.2 — Si l'état i a pour période $d(i)$, alors il existe un entier $N(i)$ (dépendant de i) tel que $\forall n \geq N(i)$,

$$P_{ii}^{nd(i)} > 0 .$$

Corollaire 4.1 — $P_{ji}^m > 0 \Rightarrow P_{ji}^{m+nd(i)} > 0 \quad \forall n$ suffisamment grand.

Définition 4.13 — Une chaîne de Markov est dite **apériodique** si tous ses états sont de période 1.

Soit un état i . On définit, pour chaque entier $n \geq 1$,

$$f_{ii}^n = \mathbb{P}(X_n = i, X_\nu \neq i, \nu = 1, 2, \dots, n-1 \mid X_0 = i)$$

qui est la probabilité que, partant de l'état i , le premier retour à cet état se passe au n^e pas de la transition.

Proposition 4.5 — Pour $n \geq 1$,

$$P_{ii}^n = \sum_{k=0}^n f_{ii}^k P_{ii}^{n-k}$$

avec $f_{ii}^0 = 0$ pour tout i .

Définition 4.14 — Un état i est dit **récurrent** ssi

$$\sum_{n=1}^{\infty} f_{ii}^n = 1 ,$$

i.e. ssi, partant de cet état, la probabilité d'y repasser après un temps fini vaut 1.

Définition 4.15 — Un état non récurrent est dit **transient**.

Théorème 4.3 — Un état i est récurrent ssi

$$\sum_{n=1}^{\infty} P_{ii}^n = \infty .$$

Corollaire 4.2 — Si $i \leftrightarrow j$ et si i est récurrent, alors j l'est aussi.

Remarque — Ceci prouve que la récurrence, comme la périodicité, est une propriété de classe : tous les états d'une même classe d'équivalence sont soit récurrents, soit transients.

Remarque — Le nombre attendu de (re)passages par l'état i , étant donné $X_0 = i$, vaut $\sum_{n=1}^{\infty} P_{ii}^n$. Par conséquent, le théorème ci-dessus dit que l'état i est récurrent ssi le nombre attendu de (re)passages par cet état est infini.

Définition 4.16 — *On définit*

$$Q_{ij} = \mathbb{P}(\text{une chaîne partant de l'état } i \text{ visite infiniment souvent l'état } j) .$$

Théorème 4.4 — *L'état } i est récurrent (respectivement transient) si } Q_{ii} = 1 (resp. } Q_{ii} = 0).*

Théorème 4.5 — *Si } i \leftrightarrow j et si la classe est récurrente, alors*

$$\begin{aligned} f^* &= \sum_{n=1}^{\infty} f_{ij}^n \\ &= 1 . \end{aligned}$$

Corollaire 4.3 — *Si } i \leftrightarrow j et si la classe est récurrente, alors } Q_{ij} = 1.*

Définition 4.17 — *Une chaîne de Markov est dite **récurrente** (respectivement **irréductible**) si tous ses états sont récurrents (resp. irréductibles).*

Théorème 4.6 (Théorème limite 1) — *Soit une chaîne de Markov récurrente, irréductible et apériodique. Soit } P_{ii}^n la probabilité de repasser en } i lors de la } n^e transition, } n = 0, 1, 2, \dots, étant donné que la chaîne part de } i, i.e. } X(0) = i. Par convention, } P_{ii}^0 = 1. Soit } f_{ii}^n la probabilité que le premier (re)passage en } i se fasse lors de la } n^e transition, } n = 0, 1, 2, \dots, avec la convention } f_{ii}^0 = 0. Alors*

$$P_{ii}^n - \sum_{k=0}^n f_{ii}^{n-k} P_{ii}^k = \begin{cases} 1 & \text{si } n = 0 , \\ 0 & \text{si } n > 0 , \end{cases}$$

et

$$\lim_{n \rightarrow \infty} \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^n} .$$

Théorème 4.7 (Théorème limite 2) — *Sous les mêmes conditions que celles du théorème précédent,*

$$\lim_{n \rightarrow \infty} P_{ji}^n = \lim_{n \rightarrow \infty} P_{ii}^n .$$

Proposition 4.6 — Si i appartient à une classe récurrente apériodique, alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^n P_{ii}^m = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^n}.$$

Proposition 4.7 — Si i appartient à une classe récurrente périodique de période d , alors $P_{ii}^m = 0$ si m n'est pas un multiple de d , et de plus

$$\lim_{n \rightarrow \infty} P_{ii}^{nd} = \frac{d}{\sum_{n=0}^{\infty} n f_{ii}^n}.$$

Remarque — $\sum_{n=0}^{\infty} n f_{ii}^n$ est le temps de récurrence moyen.

Notation — On note $\pi_i = \lim_{n \rightarrow \infty} P_{ii}^n$.

Proposition 4.8 — Si $\pi_i > 0$ pour un état i d'une classe récurrente apériodique, alors $\pi_i > 0$ pour tout état j de la classe de i . Dans ce cas, cette classe est dite **récurrente positive** ou fortement ergodique.

Proposition 4.9 — Si $\pi_i = 0$ pour tout état i d'une classe récurrente, cette classe est dite **récurrente nulle** ou faiblement ergodique.

Théorème 4.8 — Dans une classe récurrente positive et apériodique d'états $j = 0, 1, 2, \dots$,

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{jj}^n &= \pi_j \\ &= \sum_{i=0}^{\infty} \pi_i P_{ij} \end{aligned}$$

et

$$\sum_{i=0}^{\infty} \pi_i = 1.$$

Les $(\pi)_i$ sont déterminés de façon unique par les trois équations suivantes :

$$\begin{cases} \pi_i \geq 0, \\ \sum_{i=0}^{\infty} \pi_i = 1, \\ \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}. \end{cases} \quad (4.1)$$

Définition 4.18 — Tout ensemble $(\pi_i)_{i=0,1,2,\dots}$ vérifiant (4.1) est appelé **distribution de probabilité stationnaire** de la chaîne de Markov.

Définition 4.19 — Soit T l'ensemble de tous les états transients, C, C_1, C_2, \dots les classes de récurrence et i un état transient. On définit

$$\pi_i(C)$$

comme étant la probabilité que la chaîne, partant de i , soit absorbée par la classe récurrente C (rappel : une fois que la chaîne entre dans une classe récurrente, elle ne la quitte plus).

Théorème 4.9 — Soit $j \in C$ (classe récurrente apériodique). Alors pour $i \in T$,

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{ij}^n &= \pi_i(C) \cdot \lim_{n \rightarrow \infty} P_{jj}^n \\ &= \pi_i(C) \cdot \pi_j \end{aligned}$$

Théorème 4.10 — Soit \mathcal{B} une chaîne de Markov irréductible dont l'espace d'états est désigné par des entiers positifs. Une condition nécessaire et suffisante pour que \mathcal{B} soit transiente est que le système d'équations

$$\sum_{j=0}^{\infty} P_{ij} y_j = y_i, \quad i \neq 0$$

admette une solution bornée non constante.

Théorème 4.11 — Une condition suffisante pour qu'une chaîne de Markov soit récurrente qu'il existe une séquence $\{y_i\}$ telle que

$$\sum_{j=0}^{\infty} P_{ij} y_j \leq y_i \quad \text{pour } i \neq 0 \text{ et avec } y_i \rightarrow \infty.$$

5

Ergodicité

Définition 5.1 — Un processus est dit **stationnaire (au sens fort)** si, pour tout $h > 0$,

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}) \stackrel{\mathcal{L}}{=} (X_{t_1}, X_{t_2}, \dots, X_{t_n}).$$

Définition 5.2 — Un processus est dit **stationnaire par covariance (stationnaire au sens faible)** si ses moments du second ordre sont finis et si sa fonction de covariance

$$\begin{aligned} \text{Cov}(X_n, X_{n+v}) &= \mathbb{E}(X_n X_{n+v}) - \mathbb{E}(X_n) \mathbb{E}(X_{n+v}) \\ &= \mathbb{E}[(X_n - m)(X_{n+v} - m)] \\ &= R(v) \quad (\text{notation}) \end{aligned}$$

ne dépend que de h pour tout $t \in T$ — m étant la moyenne du processus.

Proposition 5.1 — Un processus stationnaire ayant ses moments d'ordre 2 finis est un processus stationnaire par covariance.

Théorème 5.1 (Ergodicité des carrés moyens) — Soit (X_n) un processus stationnaire par covariance ayant pour fonction de covariance $R(v)$. Alors

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{v=0}^{N-1} R(v) = 0 \quad \text{ssi} \quad \lim_{N \rightarrow \infty} \mathbb{E}[(\bar{X}_N - m)^2] = 0$$

où $\bar{X}_N = \frac{1}{N}(X_1 + \dots + X_N)$.

Remarque — Le théorème précédent est une généralisation de la loi des grands nombres : au lieu d'être indépendantes, les variables X_n sont **asymptotiquement indépendantes**, dans ce sens que la covariance $R(v)$ a une limite de Cesaro nulle quand v tend vers l'infini.

Théorème 5.2 — Soit (X_n) un processus gaussien stationnaire par covariance ayant pour fonction de covariance $R(v)$ et pour moyenne 0. Alors

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{v=0}^{T-1} R(v)^2 = 0 \implies \lim_{T \rightarrow \infty} \mathbb{E}[|\hat{R}_T(v) - R(v)|^2] = 0,$$

où $\hat{R}_T(v)$ est la fonction de covariance de l'échantillon, soit

$$\hat{R}_T(v) = \frac{1}{T} \sum_{l=0}^{T-1} X_l X_{l+v}.$$

Théorème 5.3 (Ergodicité des carrés moyens) — Soit (X_n) un processus stationnaire par covariance. Alors il existe une variable aléatoire \bar{X} telle que

$$\lim_{n \rightarrow \infty} \|\bar{X}_N - \bar{X}\|_2 = 0.$$

Théorème 5.4 (Ergodicité des carrés moyens) — Soit (X_n) un processus (faiblement) stationnaire de moyenne $\mathbb{E}(X_n) = m$. Alors $\bar{X}_n = \frac{1}{n}(X_0 + \dots + X_{n-1})$ converge en probabilité vers une variable aléatoire \hat{X} , ce qui s'écrit

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \bar{X}_n = \bar{X}\right) = 1.$$

Définition 5.3 — On appelle **opérateur de shift** l'opérateur T défini par :

$$\begin{aligned} Tx &= T(x_0, x_1, x_2, \dots) \\ &= (x_1, x_2, x_3, \dots). \end{aligned}$$

Définition 5.4 — On appelle **ensemble invariant par opération de shift** un ensemble A tel que, si T est l'opérateur de shift, alors

$$Tx \text{ est un élément de } A \iff x \text{ est dans } A.$$

Définition 5.5 — Soit (X_n) un processus (faiblement) stationnaire. Il est dit **ergodique** si, pour tout ensemble A invariant par opération de shift,

$$\mathbb{P}((X_0, X_1, \dots) \in A) = 0 \text{ ou } 1.$$

Théorème 5.5 — Soit (X_n) un processus stationnaire ergodique de moyenne finie $\mathbb{E}(X_n) = m$. Alors, avec une probabilité 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) = m.$$

Théorème 5.6 — Soit (X_n) un processus stationnaire. Les conditions suivantes sont équivalentes :

- (i) (X_n) est ergodique ;
- (ii) pour tout ensemble A invariant par opération de shift,

$$\mathbb{P}((X_0, X_1, \dots) \in A) = 0 \text{ ou } 1 ;$$

- (iii) pour tout ensemble A de la forme (x_0, x_1, \dots) ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{(X_j, X_{j+1}, \dots) \in A\}} = \mathbb{P}((X_0, X_1, \dots) \in A) ;$$

- (iv) pour tout $k = 1, 2, \dots$ et tout ensemble A de la forme (x_0, \dots, x_k) ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{(X_j, \dots, X_{j+k}) \in A\}} = \mathbb{P}((X_0, \dots, X_k) \in A) ;$$

- (v) pour tout k et toute fonction ϕ de $k+1$ variables,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \phi(X_j, \dots, X_{j+k}) = \mathbb{E}[\phi(X_0, \dots, X_k)] ;$$

à condition que cette espérance existe ;

- (vi) pour toute fonction ϕ sur un ensemble (x_0, \dots, x_k) ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \phi(X_j, X_{j+1}, \dots) = \mathbb{E}[\phi(X_0, X_1, \dots)] ,$$

à condition que cette espérance existe.

6

Entropie

Tandis qu'une probabilité mesure l'incertitude touchant l'occurrence d'un événement, l'**entropie** mesure l'incertitude touchant l'occurrence d'un ensemble d'événements.

Définition 6.1 — Soit X une v.a. prenant la valeur i avec la probabilité p_i , $i = 1, \dots, n$. L'**entropie** de X se définit par

$$H(X) = - \sum_{i=1}^n p_i \log(p_i)$$

(avec la convention $0 \times \log 0 = 0$).

Propriété 6.1 — L'entropie vérifie les trois propriétés suivantes :

- (i) l'entropie d'une variable aléatoire constante est nulle ;
- (ii) l'ajout à l'entropie de la valeur $i + 1$, avec la probabilité correspondante p_{i+1} , ne modifie pas l'entropie ;
- (iii) l'entropie est maximisée, avec la valeur maximum $\log n$, lorsque $p_1 = \dots = p_n = 1/n$.

Remarque — La propriété (iii) est conforme à l'intuition, qui veut que la v.a. X_1 prenant les valeurs 0 et 1 avec les probabilités 0,001 et 0,999 est plus prévisible que la v.a. X_2 prenant les valeurs 0 et 1 avec probabilité $1/2$.

Définition 6.2 — On définit l'entropie d'un couple de v.a. (X, Y) par

$$H(X, Y) = - \sum_{i, j} p_{ij} \log(p_{ij}) .$$

Définition 6.3 — On définit l'entropie conditionnelle de X sachant Y par

$$H(X | Y) = - \sum_j \mathbb{P}(Y = j) \sum_i p(i | j) \log [p(i | j)]$$

(avec $p(i | j) = \mathbb{P}(X = i | Y = j)$).

Proposition 6.1 — $H(X | Y) = H(X, Y) - H(Y)$.

Proposition 6.2 — $H(X_k | X_1, \dots, X_{k-1}) \leq H(X_k | X_2, \dots, X_{k-1})$.

Définition 6.4 — On définit l'entropie d'un processus (X_n) par

$$H[(X_n)] = \lim_{k \rightarrow \infty} H(X_k | X_1, \dots, X_{k-1}).$$

Proposition 6.3 — $H[(X_n)] = \lim_{k \rightarrow \infty} \frac{1}{k} H(X_1, \dots, X_k)$.

Proposition 6.4 — Si (X_n) est ergodique, alors

$$H[(X_n)] = \lim_{n \rightarrow \infty} -\frac{1}{n} \log p(X_0, \dots, X_{n-1}).$$

Proposition 6.5 — Soit (X_n) une chaîne de Markov irréductible d'espace d'états fini. On suppose que $\pi(i) = \mathbb{P}(X_0 = i)$, où $(\pi(i))_{i=1, \dots, N}$ est la distribution stationnaire de la chaîne. Alors

$$H[(X_n)] = - \sum_{i,j} \pi(i) P(i, j) \log P(i, j).$$

Proposition 6.6 — Une chaîne de Markov irréductible d'espace d'états fini commençant avec sa distribution stationnaire est un processus stationnaire ergodique.

Théorème 6.1 — Soit (X_n) un processus stationnaire ergodique d'espace d'états fini $\{1, \dots, N\}$. Soient

$$p(i_1, \dots, i_m) = \mathbb{P}(X_1 = i_1, \dots, X_m = i_m)$$

et

$$H[(X_n)] = \lim_{n \rightarrow \infty} \left(-\frac{1}{n} \sum_{k=1}^n \sum_{i_1, \dots, i_k} p(i_1, \dots, i_k) \log p(i_1, \dots, i_k) \right).$$

Alors, avec une probabilité 1,

$$H[(X_n)] = \lim_{n \rightarrow \infty} \left[-\frac{1}{n} \log p(X_1, \dots, X_n) \right].$$

Troisième partie

**PROCESSUS
STOCHASTIQUES**

Généralités

Définition 7.1 — On considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ où \mathbb{P} est la mesure de probabilité sur (Ω, \mathcal{A}) . Un **processus aléatoire**, ou encore une **fonction aléatoire réelle** (f.a.r.) est une fonction à deux variables : t — le temps — et ω — le hasard —, et elle est notée $X(t, \omega)$, avec $t \in [0, \infty[$ et $\omega \in \Omega$.

À t fixé, la fonction $X_t : \omega \mapsto X(t, \omega)$ est appelée **coordonnée** à l'instant t (c'est donc une v.a.). La **trajectoire** est $\omega \mapsto (X(t, \omega), t \geq 0)$, ordinairement continue.

Une f.a.r. à trajectoire continue (f.a.r.c.) est une application

$$\begin{aligned} X : [0, \infty[\times \Omega &\rightarrow \mathbb{R} \\ (t, \omega) &\mapsto X(t, \omega) \end{aligned}$$

telle que :

- a) pour presque tout ω , $t \mapsto X(t, \omega)$ est continue ;
- b) pour tout $t \geq 0$, $X_t : \omega \mapsto X(t, \omega)$ est une v.a.r.

La loi de X est caractérisée par la loi des $(X_{t_1}, \dots, X_{t_k})_{k \geq 1, 0 \leq t_1 < \dots < t_k < \infty}$. En fait, il s'agit d'une loi marginale finie k -dimensionnelle. Soit

$$\begin{aligned} X : \Omega &\rightarrow \mathcal{C}(\mathbb{R}^+, \mathbb{R}) \\ \omega &\mapsto (X(t, \omega), t \geq 0) \end{aligned}$$

où $\mathcal{C}(\mathbb{R}^+, \mathbb{R})$ est munie de la topologie de la convergence uniforme sur les compacts. Munissant $\mathcal{C}(\mathbb{R}^+, \mathbb{R})$ de la tribu borélienne, X est mesurable. Par conséquent, l'image de \mathbb{P} par cette application mesurable est la probabilité sur \mathcal{C} notée \mathbb{P}_X .

7.1 Espaces gaussiens

Définition 7.2 — Un sev¹ fermé F de $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ est un **espace gaussien** si ses éléments sont des v.a. gaussiennes centrées. Étant donné X une f.a.r. gaussienne, on note $H^X = \overline{\text{vect}(X - \mathbb{E}(X))}^{\mathcal{L}^2}$ l'espace gaussien associé à X .

$(H^X, \langle \cdot, \cdot \rangle)$ est un espace de Hilbert (car c'est un fermé inclus dans un complet, donc il est complet). Si X est continue, alors H^X est séparable. Soit $(\zeta_n)_n$ une base orthonormée de H^X . Développons $X_t - \mathbb{E}(X_t)$ sur cette base (formule de **Karhunen – Loeve**) :

$$X(t) = \mathbb{E}(X(t)) + \sum_n c_n(t) \zeta_n(\omega)$$

avec

$$\begin{aligned} c_n(t) &= \langle X_t - \mathbb{E}(X(t)), \zeta_n \rangle \\ &= \mathbb{E}(\zeta_n [X_t - \mathbb{E}(X_t)]) \\ &= \mathbb{E}(\zeta_n X_t) . \end{aligned}$$

7.2 Mouvement brownien

Définition 7.3 — Un **mouvement brownien** est une f.a.r.c. $B(t, \omega)$ à accroissements indépendants gaussiens, avec :

- (i) $B(t) - B(s) \rightsquigarrow \mathcal{N}(0, t - s)$ pour $0 \leq s < t$;
- (ii) $B(0) = 0$.

Propriété 7.1 — B est un processus gaussien centré à trajectoire continue et de covariance

$$\mathbb{E}(B(s)B(t)) = \min(s, t) .$$

La réciproque est vraie.

Propriété 7.2 — Si B est un mouvement brownien, il en est de même de

$$X(t) = \frac{1}{c} B(c^2 t)$$

et

$$Y(t) = t B\left(\frac{1}{t}\right) ,$$

pour $c \in \mathbb{R}^*$.

1. Sous-espace vectoriel.

7.3 Principe d'invariance

Soit $(\zeta_n)_n$ des v.a. i.i.d. d'espérance nulle et de variance σ^2 finie. Soit $S_n = \sum_i \zeta_i$. D'après le théorème de la limite centrale,

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) .$$

Soit la **marche aléatoire renormalisée**

$$X_t^{(n)} = \frac{\sum_{i=1}^{[nt]} \zeta_i + (nt - [nt])\zeta_{[nt]+1}}{\sigma\sqrt{n}} .$$

Théorème 7.1 (Donsker) — *La suite de processus $X_t^{(n)}$ converge en loi vers B quand n tend vers l'infini.*

7.4 Propriétés du brownien

7.4.1 Variation quadratique

Nous savons que :

- (i) $B(t+h) - B(t) \rightsquigarrow \mathcal{N}(0, h)$;
- (ii) $(B(t+h) - B(t))/h \rightsquigarrow \mathcal{N}(0, h \times (1/h^2) = 1/h)$, qui n'a pas de limite quand h décroît vers 0.

Par conséquent, nous avons le résultat suivant.

Proposition 7.1 — *Le brownien n'est pas dérivable.*

Partitionnons $[0, T]$: $0 = t_0 < t_1 < \dots < t_n = T$. Le **pas** de cette partition est

$$\Delta = \max_{1 \leq i \leq n} (t_i - t_{i-1}) .$$

Définition 7.4 — *La **variation totale** est définie comme étant*

$$\lim_{\Delta \rightarrow 0} \sum_i |B(t_i) - B(t_{i-1})| .$$

et elle est infinie.

Définition 7.5 — *La **variation quadratique** est*

$$\lim_{\Delta \rightarrow 0} \sum_i |B(t_i) - B(t_{i-1})|^2$$

et elle est finie. On la note $\langle B_t \rangle$ ou $\langle B \rangle_t$.

Théorème 7.2 — La variation quadratique de B existe dans \mathcal{L}^2 et vaut p.s. T .

Proposition 7.2 — Pour presque tout ω ,

- (i) $t \rightarrow B(t, \omega)$ n'est dérivable en aucun t ;
- (ii) $\forall \alpha < 1/2$, $|B(t) - B(s)| \leq c|t - s|^\alpha$, quels que soient s et t dans $[0, T]$ — par ailleurs, $c = c(\alpha, \omega, T) < \infty$.

7.4.2 Martingales

Définition 7.6 — Soit $(\Omega, \mathcal{A}, \mathbb{P})$, $t \in \mathbb{N}$ ou \mathbb{R}^+ . Une **filtration** est une famille \mathcal{F}_t de tribus, $t \in \mathbb{N}$ ou \mathbb{R}^+ , telle que

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{A}$$

$\forall s \leq t$.

Définition 7.7 — Soient $(\Omega, \mathcal{A}, \mathbb{P})$ et \mathcal{F}_t une filtration. Un processus $X = X(t, \omega)$ est dit **\mathcal{F}_t -adapté** si $\forall t$, X_t est \mathcal{F}_t -mesurable.

Définition 7.8 — Soit $(\Omega, \mathcal{A}, \mathbb{P})$. Soit $(M_t)_t$, $t \in \mathbb{N}$ ou \mathbb{R}^+ , un processus réel défini sur Ω . Soit $(\mathcal{F}_t)_t$ une filtration sur Ω . $(M_t)_t$ est une **\mathcal{F}_t -martingale** si :

- (i) $\forall t$, M_t est \mathcal{F}_t -adaptée et $M_t \in \mathcal{L}^1$;
- (ii) pour $0 \leq s \leq t$, $\mathbb{E}(M_t | \mathcal{F}_s) = M_s$ p.s.

Conséquence — $\mathbb{E}(M_t) = \mathbb{E}(M_s) = \mathbb{E}(M_0)$.

Exemples — Nous donnons quelques exemples de martingales.

- 1) **Marche aléatoire** : soit $\mathcal{F}_n = \sigma(\zeta_1, \dots, \zeta_n)$. Soit $S_n = \sum_{i=1}^n \zeta_i$, avec les ζ_i i.i.d. centrées. Alors S_n est une \mathcal{F}_n -martingale : $\zeta_i \in \mathcal{L}^1 \Rightarrow S_n \in \mathcal{L}^1$. S_n est \mathcal{F}_n -mesurable.

$$\begin{aligned} \mathbb{E}(S_n | \mathcal{F}_{n-1}) &= \mathbb{E}(S_{n-1} + \zeta_n | \mathcal{F}_{n-1}) \\ &= S_{n-1} \end{aligned}$$

$$\begin{aligned} \mathbb{E}(S_n | \mathcal{F}_{n-2}) &= \mathbb{E}[\mathbb{E}(S_n | \mathcal{F}_{n-1}) | \mathcal{F}_{n-2}] \\ &= \mathbb{E}(S_{n-1} | \mathcal{F}_{n-2}) \\ &= S_{n-2} . \end{aligned}$$

- 2) **Brownien** : soit $\mathcal{F}_t = \sigma(B_u, u \leq t)$. B est une \mathcal{F}_t -martingale ; soit $s \leq t$:

$$\begin{aligned} \mathbb{E}(B_t | \mathcal{F}_s) &= \mathbb{E}(B_t - B_s + B_s | \mathcal{F}_s) \\ &= \mathbb{E}(B_t - B_s | \mathcal{F}_s) + B_s \\ &= \mathbb{E}(B_t - B_s) + B_s \\ &= B_s , \end{aligned}$$

car $(B_t - B_s) \perp B_u$, $u \leq s$.

3) Soit $\mathcal{F}_n = \sigma(\zeta_1, \dots, \zeta_n)$. Soit $S_n = \sum_{i=1}^n \zeta_i$, avec les ζ_i i.i.d. centrées de variance σ^2 . Alors $X_n = S_n - n\sigma^2$ est une \mathcal{F}_n -martingale.

4) Soit $\mathcal{F}_t = \sigma(B_u, u \leq t)$. Alors

$$M_t = B_t^2 - t$$

est une \mathcal{F}_t -martingale.

5) Soit $\mathcal{F}_t = \sigma(B_u, u \leq t)$. Alors

$$Y_t = \exp \left[\lambda B_t - \frac{\lambda^2 t}{2} \right],$$

pour $\lambda \in \mathbb{C}$, est une \mathcal{F}_t -martingale.

Remarque — Soit $X(t)$ une f.a.r.c. telle que $X(0) = 0$ et telle que $\exp \left[\lambda X_t - \frac{\lambda^2 t}{2} \right]$ soit une \mathcal{F}_t -martingale, avec $\mathcal{F}_t = \sigma(X_s, s \leq t)$, $\lambda \in \mathbb{R}$ (ou $\lambda \in i\mathbb{R}$). Alors X est un brownien.

Définition 7.9 — X est une **\mathcal{F}_t -sous-martingale** (respectivement une **\mathcal{F}_t -sur-martingale**) si :

(i) $\forall t, X_t$ est \mathcal{F}_t -mesurable ;

(ii) $\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s$ p.s. (resp. $\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$ p.s.), $\forall 0 \leq s \leq t$.

Proposition 7.3 — Soit M une martingale et $\phi : \mathbb{R} \rightarrow \mathbb{R}$ convexe. Alors $X_t = \phi(M_t)$ est une sous-martingale.

En particulier, le résultat précédent, pour $\phi(x) = x^2$, nous indique que le carré d'une martingale est une martingale.

Proposition 7.4 — Soit M une martingale continue et $\langle M \rangle_t$ sa variation quadratique. Alors

$$X_t = M_t^2 - \langle M \rangle_t .$$

est une martingale.

7.4.3 Théorème d'arrêt — Inégalité de Doob

Définition 7.10 — Soient $(\Omega, \mathcal{A}, \mathbb{P})$ et $(\mathcal{F}_t)_{t \geq 0}$ une filtration. On appelle **temps d'arrêt** une v.a. $T : \Omega \rightarrow [0, \infty]$ telle que $\forall t \geq 0, \{T \leq t\} \in \mathcal{F}_t$.

Exemples — Voici quelques exemples.

1) $T = t_0, \forall \omega$, avec $t_0 \geq 0$, est un temps d'arrêt.

2) X processus \mathcal{F}_t -adapté, $A \in \mathcal{B}(\mathbb{R})$. Le **temps d'entrée** dans A est

$$T_A = \inf \{ t \geq 0, X_t \in A \} .$$

- 3) Si A est un ouvert et si X est continue à droite, alors T_A est un \mathcal{F}_{t+} -temps d'arrêt, où $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$.
- 4) Si B est un fermé et si X est continue, alors T_B est un \mathcal{F}_t -temps d'arrêt.
- 5) $T \wedge T'$ est un temps d'arrêt si T et T' en sont.

Théorème 7.3 (Théorème d'arrêt) — Soient $(M_t)_t$ une \mathcal{F}_t -martingale et T un \mathcal{F}_t -temps d'arrêt p.s. borné (i.e. $T \leq \text{cte}$ p.s.). On suppose que $(M_t)_t$ est continue à droite. Soit $0 \leq s \leq T$. Alors

$$\mathbb{E}(M_T | \mathcal{F}_s) = M_s \text{ p.s.}$$

Le résultat demeure :

- pour une sous-martingale, avec la relation $\mathbb{E}(M_T | \mathcal{F}_s) \geq M_s$ p.s. ;
- pour une sur-martingale, avec la relation $\mathbb{E}(M_T | \mathcal{F}_s) \leq M_s$ p.s.

Corollaire 7.1 — Nous avons :

$$\mathbb{E}(M_T) = \mathbb{E}(M_s) = \mathbb{E}(M_0) .$$

Proposition 7.5 (Inégalité de Doob) — Soit X_t une \mathcal{F}_t -martingale de carré intégrable et continue à droite. Alors $\forall T \geq 0, \forall \lambda > 0$,

$$\mathbb{P}\left(\sup_{0 \leq t \leq T} |X_t| > \lambda\right) \leq \frac{1}{\lambda^2} \mathbb{E}(X_T^2) .$$

Théorème 7.4 — Nous avons :

$$\lim_{t \rightarrow \infty} \frac{B_t}{t} = 0 \text{ p.s.}$$

Proposition 7.6 — Soient $s \geq 0$ et $X(t) = B(t+s) - B(s)$. Alors X est un brownien et est indépendant de $\mathcal{F}_s = \sigma(B_u, u \leq s)$.

Proposition 7.7 — Si T est un temps d'arrêt pour la filtration du brownien, alors $X(t) = B(t+T) - B(T)$ est aussi un brownien.

7.4.4 Intégrale de Wiener

$\mathcal{L}^2(\Omega)$ est muni du p.s. produit scalaire $\mathbb{E}(XY)$. $\mathcal{L}^2(\mathbb{R}^+)$ est muni du p.s. produit scalaire $\langle f, g \rangle = \int_{\mathbb{R}^+} f(t)g(t) dt$. On veut définir

$$\int_{\mathbb{R}^+} \mathbf{1}_{]u, v]}(s) dB(s) = B(v) - B(u)$$

pour $0 \leq u < v$.

Définition 7.11 — Soient $0 = t_0 \leq t_1 < \dots < t_n$. Alors

$$\int_{\mathbb{R}^+} \left\{ \sum_{k=1}^n a_k \mathbf{1}_{]t_{k-1}, t_k]}(s) \right\} dB(s) = \sum_{k=1}^n a_k (B(t_k) - B(t_{k-1})) \quad (7.1)$$

sur $F = \{ \text{fonctions en escalier} : f(s) = \sum_{k=1}^n a_k \mathbf{1}_{]t_{k-1}, t_k]}(s) \}$.

Remarque — Cette définition ne dépend pas du choix de la fonction étagée.

L'équation (7.1) s'écrit Tf , avec $T : F \rightarrow H^B$ (espace gaussien engendré par le brownien); $F \subset \mathcal{L}^2(\mathbb{R})$, $H^B \subset \mathcal{L}^2(\Omega)$. Cette application linéaire est de norme 1.

Théorème 7.5 — Soit B un brownien sur $(\Omega, \mathcal{A}, \mathbb{P})$. On peut associer à toute fonction f de $\mathcal{L}^2(\mathbb{R})$ une v.a. centrée, intégrable et gaussienne de H^B , notée $\int_{\mathbb{R}^+} f(t) dB(t)$ et appelée **intégrale de Wiener** telle que :

- (i) $\int_{\mathbb{R}^+} \mathbf{1}_{]u, v]}(t) dB(t) = B(v) - B(u)$ pour $u < v$;
- (ii) $f \mapsto \int_{\mathbb{R}^+} f(t) dB(t)$ est linéaire et isométrique;
- (iii) on a les propriétés suivantes :

$$\begin{aligned} \mathbb{E} \left[\int_{\mathbb{R}^+} f(t) dB(t) \times \int_{\mathbb{R}^+} g(t) dB(t) \right] &= \int_{\mathbb{R}^+} f(t) g(t) dt . \\ \mathbb{E} \left[\int_{\mathbb{R}^+} f(t) dB(t) \times B(s) \right] &= \int_0^s f(t) dt , \\ \left\{ \int_{\mathbb{R}^+} f dB , f \in \mathcal{L}^2(\mathbb{R}^+) \right\} &= H^B . \end{aligned}$$

Proposition 7.8 (Intégration par parties) — Si $f \in \mathcal{C}^1(\mathbb{R}^+)$, alors p.s.

$$\int_0^t f(s) dB(s) = f(t) B(t) - \int_0^t f'(s) B(s) ds .$$

7.4.5 Équation de Langevin

A. Cas unidimensionnel

Définition 7.12 — *L'équation de Langevin est*

$$dV(t) = -bV(t) dt + \sigma dB(t)$$

ce qui équivaut à

$$V(t) = V(0) - \int_0^t bV(s) ds + \sigma B(t).$$

Proposition 7.9 — *La solution de l'équation de Langevin est donnée par*

$$V(t) = e^{-bt} V(0) + \int_0^t \sigma e^{-b(t-s)} dB(s).$$

Définition 7.13 — *V s'appelle le processus d'Ornstein-Uhlenbeck.*

Proposition 7.10 — *On suppose que $V(0)$ est indépendant de B et qu'il suit une loi normale centrée de variance $\sigma^2/(2b)$. Alors $V(t)$ est un processus gaussien stationnaire.*

B. Cas multidimensionnel

Soit

$$dV(t) = -bV(t) dt + \sigma dB(t)$$

avec $V \in \mathbb{R}^d$, $b \in \mathcal{M}_{d \times d}$, $\sigma \in \mathcal{M}_{d \times d}$, $b \in \mathbb{R}^d$ et $B = (B_1, \dots, B_d)^t$ brownien de dimension d .

Proposition 7.11 — *On suppose que :*

- (i) $V(0)$ est indépendant de B ;
- (ii) $\mathbb{E}(V(0)) = 0$;
- (iii) $\mathbb{V}(V(0)) = \sigma^2/(2b)$;
- (iv) $V(0) \rightsquigarrow \mathcal{N}(0, \sigma^2/(2b))$.

Alors $V(t)$ est un processus gaussien stationnaire.

Calcul stochastique

8.1 Intégrale stochastique d'Ito

8.1.1 Filtration

Définition 8.1 — Une f.a. $\phi(t, \omega)$ définie sur $\mathbb{R}^+ \times \Omega$ (respectivement sur $[0, T] \times \Omega$) est dite **progressivement mesurable** par rapport à la filtration $\mathcal{F} = (\mathcal{F}_t, t \geq 0)$ si $\forall t \in \mathbb{R}^+$ (resp. $t \leq T$), la restriction de ϕ suivante :

$$\begin{aligned} \phi : [0, t] \times \Omega &\rightarrow \mathbb{R} \\ (s, \omega) &\mapsto \phi(s, \omega) \end{aligned}$$

est mesurable par rapport à $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$.

Remarque — $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ est engendré par les $B \times A$, $\forall B \in \mathcal{B}([0, t])$, $\forall A \in \mathcal{F}_t$.

On note $M^2(\mathbb{R}^+)$ (resp. $M^2([0, t])$) l'ensemble des fonctions ϕ progressivement mesurables et telles que

$$\mathbb{E} \left[\int_{\mathbb{R}^+} \phi^2(t, \omega) dt \right] < \infty .$$

(resp. intégrale sur $[0, T]$), et l'on considèrera par la suite

$$M^2 = \bigcap_T M^2([0, t]) .$$

Nota — Soient ϕ progressivement mesurable et t fixé. $\omega \mapsto \phi(t, \omega)$ est \mathcal{F}_t -mesurable, donc ϕ est adaptée.

Par la suite, nous travaillerons sur \mathbb{R}^+ , mais les résultats seront également valables pour $M^2([0, t])$.

$M^2(\mathbb{R}^+)$ est un espace de Hilbert : on a le produit scalaire

$$\langle \phi, \psi \rangle_{M^2(\mathbb{R}^+)} = \mathbb{E} \left[\int_{\mathbb{R}^+} \phi^2(t, \omega) \psi(t, \omega) \right] dt .$$

8.1.2 Fonctions en escalier

Soit $0 = t_0 < t_1 < \dots < t_n$. On définit

$$\phi(t, \omega) = \sum_{i=0}^{n-1} X_i(\omega) \mathbf{1}_{]t_i, t_{i+1}]}(t) ,$$

avec X_i \mathcal{F}_{t_i} -mesurable et $X_i \in \mathcal{L}^2(\mathcal{F}_{t_i})$.

$$\left(\phi|_{[0, T] \times \Omega} \right)^{-1} (I) = \bigcup_{i : t_i \leq t} X_i^{-1}(I) \times (]t_i, t_{i+1}] \cap [0, t])$$

appartient à $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$. Par conséquent, ϕ est mesurable.

Définition 8.2 — Pour ϕ en escalier, on définit

$$\int_{\mathbb{R}^+} \phi(t) dB(t) = \sum_{i=0}^{n-1} X_i [B(t_{i+1}) - B(t_i)] .$$

Proposition 8.1 — Nous avons :

$$\mathbb{E} \left[\int_{\mathbb{R}^+} \phi(t) dB(t) \right] = \sum_{i=0}^{n-1} \mathbb{E} [X_i (B(t_{i+1}) - B(t_i))] .$$

et

$$\mathbb{E} \left[\int_{\mathbb{R}^+} \phi(t) dB(t) \right]^2 = \sum_{i=0}^{n-1} \mathbb{E} (X_i^2) (t_{i+1} - t_i) .$$

Corollaire 8.1 — L'intégrale stochastique est une isométrie.

En effet,

$$\mathbb{E} \left[\int_{\mathbb{R}^+} \phi(t) dB(t) \right]^2 = \mathbb{E} \left[\int_{\mathbb{R}^+} \phi^2(t) dt \right] .$$

L'isométrie est (IS) : $M^2(\mathbb{R}^+) \rightarrow \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ et

$$\| \text{IS}(\phi) \|_{\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})} = \|\phi\|_{M^2(\mathbb{R}^+)} .$$

8.1.3 Densité des fonctions en escaliers dans $M^2(\mathbb{R}^+)$

Soit le processus d'approximation

$$P_n : \mathcal{L}^2(\mathbb{R}^+) \rightarrow \mathcal{L}^2(\mathbb{R}^+)$$

$$f \mapsto P_n f(t) = n \sum_{i=1}^{n^2} \left(\int_{\frac{i-1}{n}}^{\frac{i}{n}} f(s) ds \right) \mathbf{1}_{] \frac{i-1}{n}, \frac{i}{n}]}(t) .$$

Alors

$$\|P_n f\|_2 \leq \|f\|_2 ,$$

$$P_n f \xrightarrow{\mathcal{L}^2(\mathbb{R}^+)} f, \forall f \in \mathcal{L}^2(\mathbb{R}^+)$$

et

$$P_n \phi \xrightarrow{\mathcal{M}^2(\mathbb{R}^+)} \phi, \forall \phi \in \mathcal{M}^2(\mathbb{R}^+) .$$

8.1.4 Intégrale stochastique

On prolonge (IS) à $\mathcal{M}^2(\mathbb{R}^+)$ par

$$\int \phi(t) dB(t) = \lim_{n \rightarrow \infty} \int P_n \phi(t) dB(t)$$

$$= \lim_{n \rightarrow \infty} n \sum_{i=1}^{n^2} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \phi(s, \omega) ds [B(t_{i+1}) - B(t_i)] .$$

qui appartient à $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. On a :

$$\mathbb{E} \left[\int_{\mathbb{R}^+} \phi(t) dB(t) \right] = 0 ,$$

$$\mathbb{E} \left[\int_{\mathbb{R}^+} \phi(t) dB(t) \right]^2 = \mathbb{E} \left[\int_{\mathbb{R}^+} \phi^2(t) dt \right]$$

et

$$\mathbb{E} \left[\left\{ \int_{\mathbb{R}^+} \phi(t) dB(t) \right\} \left\{ \int_{\mathbb{R}^+} \psi(t) dB(t) \right\} \right] = \mathbb{E} \left[\int_{\mathbb{R}^+} \phi(t) \psi(t) dt \right] .$$

8.2 L'intégrale stochastique comme martingale

Proposition 8.2 — *L'application*

$$t \rightarrow \int_0^t \phi(s) dB(s)$$

est continue en moyenne quadratique p.s.

Proposition 8.3 —

$$X(t) = \int_0^t \phi(s) dB(s)$$

est une \mathcal{F} -martingale.

Proposition 8.4 —

$$X^2(t) - \int_0^t \phi^2(s) ds$$

est une \mathcal{F} -martingale.

8.3 Formule d'Ito

8.3.1 Introduction

Première formule d'Ito

$$B^2(t) = 2 \int_0^t B(s) dB(s) + t .$$

Formule d'Ito pour les fonctions C_b^2

Soit $C_b^2 = \{\Phi : \mathbb{R} \rightarrow \mathbb{R}, C^2, \text{ avec } \Phi, \Phi', \Phi'' \text{ bornées}\}$.

Proposition 8.5 — Pour $\Phi \in C_b^2$, on a \mathbb{P} -p.s.

$$\Phi(B(t)) = \Phi(B(0)) + \int_0^t \Phi'(B(s)) dB(s) + \frac{1}{2} \int_0^t \Phi''(B(s)) ds .$$

On utilisera la relation différentielle

$$d\Phi(B(t)) = \Phi'(B(t)) dB(t) + \frac{1}{2} \Phi''(B(t)) dt .$$

8.3.2 Formule générale

Soit

$$X(t) = X(0) + \int_0^t \phi(s) dB(s) + \int_0^t \psi(s) ds ,$$

avec $\phi, \psi \in M^2$ et $X(0) \in \mathcal{L}^2(\mathcal{F}_0)$.

Proposition 8.6 (Formule générale d'Ito) — Pour $\Phi \in C_b^2$,

$$\begin{aligned}\Phi(X(t)) &= \Phi(X(0)) + \int_0^t \Phi'(X(s)) \phi(s) dB(s) + \int_0^t \Phi'(X(s)) \psi(s) ds \\ &\quad + \frac{1}{2} \int_0^t \Phi''(X(s)) \phi^2(s) ds .\end{aligned}$$

Cette formule s'écrit :

$$d\Phi(X(t)) = \Phi'(X(t)) dX(t) + \frac{1}{2} \Phi''(X(t)) \phi^2(t) dt .$$

ou encore

$$d\Phi(X) = \Phi'(X) dX + \frac{1}{2} \Phi''(X) \langle dX, dX \rangle ,$$

avec

$$\langle \phi dB + \psi dt, \phi dB + \psi dt \rangle = \phi^2 \langle dB, dB \rangle + 2\phi\psi \langle dB, dt \rangle + \psi^2 \langle dt, dt \rangle$$

où

$$\begin{cases} \langle dB, dB \rangle = dt, \\ \langle dB, dt \rangle = 0, \\ \langle dt, dt \rangle = 0. \end{cases}$$

Exemple — Soit

$$M(t) = \exp \left\{ \int_0^t \phi(s) dB(s) - \frac{1}{2} \int_0^t \phi^2(s) ds \right\} .$$

C'est une martingale.

8.3.3 Localisation

Définition 8.3 — $\phi \in M_{loc}^2([0, T])$ si :

- (i) ϕ est progressivement mesurable ;
- (ii) $\int_0^T \phi^2(t) dt < \infty$ p.s.

Nous définissons

$$M_{loc}^2 = \bigcap_{T \geq 0} M_{loc}^2([0, T]) .$$

Définition 8.4 — Si $\phi \in M_{loc}^2$, on définit le temps d'arrêt par :

$$\tau_n = \begin{cases} \inf \left\{ t \geq 0 : \int_0^t \phi^2(s) ds \geq n \right\}, \\ +\infty \text{ si } \left\{ t \geq 0 : \int_0^t \phi^2(s) ds \geq n \right\} = \emptyset. \end{cases}$$

Proposition 8.7 — Si n croît vers $+\infty$, alors τ_n croît vers $+\infty$ p.s. Par ailleurs, $\mathbf{1}_{[0, \tau_n]}(t) \phi(t) \in M^2$.

On peut alors définir

$$\int_0^t \mathbf{1}_{[0, \tau_n]}(s) \phi(s) dB(s) .$$

On vérifie que cette intégrale converge p.s. quand $n \rightarrow \infty$. On définit

$$\int_0^t \phi(s) dB(s) \stackrel{\text{p.s.}}{=} \lim \int_0^t \mathbf{1}_{[0, \tau_n]}(s) \phi(s) dB(s) .$$

pour $\phi \in M_{\text{loc}}^2$.

Définition 8.5 — $X(t) = \int_0^t \phi(s) dB(s)$, avec $\phi \in M_{\text{loc}}^2$, est une **martingale locale** s'il existe une suite $(\tau_n)_n$ de \mathcal{F} -temps d'arrêt telle que :

- (i) τ_n croît vers $+\infty$ p.s. ;
- (ii) $Y_n(t) = X(t \wedge \tau_n)$ est une \mathcal{F} -martingale pour tout n .

Remarque — $X(t)$ n'est pas forcément intégrable.

8.3.4 Cas vectoriel

Soit B un brownien dans \mathbb{R}^k , $\phi_{ij} \in M_{\text{loc}}^2$, $1 \leq i \leq d$, $1 \leq j \leq k$.

$$\left(\sum_{j=1}^k \int_0^t \phi_{ij}(s) dB_j(s) \right)_{i \leq d} = \int_0^t \phi(s) dB(s) .$$

Nous avons que $\phi \in (M_{\text{loc}}^2)^{d \times k}$ et

$$\mathbb{E} \left[\int_0^t \phi(s) dB(s) \right] = 0 ,$$

$$\mathbb{E} \left[\left(\int_0^t \phi(s) dB(s) \right) \left(\int_0^t \psi(s) dB(s) \right)^t \right] = \mathbb{E} \left[\int_0^t \phi(s) \psi(s)^t ds \right] ,$$

$$\mathbb{E} \left[\left(\int_0^t \phi(s) dB(s) \right)^t \left(\int_0^t \psi(s) dB(s) \right) \right] = \mathbb{E} \left[\int_0^t \text{tr}(\phi(s) \psi(s)^t) ds \right] .$$

Soit $\Phi \in C^{1,2}$:

$$\begin{aligned} \Phi &: \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathbb{R} \\ (t, B_1(t), \dots, B_d(t)) &\mapsto \Phi(t, B_1(t), \dots, B_d(t)) . \end{aligned}$$

Soit $X \in \mathbb{R}^d$:

$$dX(t) = \underbrace{\psi(t)}_{\in \mathbb{R}^d} dt + \underbrace{\phi(t)}_{\in \mathcal{M}_{d \times k}} dB(t) .$$

Formule d'Ito

$$\begin{aligned} \Phi(t, X_t) &= \Phi(0, X_0) + \int_0^t \frac{\partial}{\partial t} \Phi(s, X_s) ds + \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x_i} \Phi(s, X_s) \psi_i(s) ds \\ &+ \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x_i} \Phi(s, X_s) \sum_{j=1}^k \phi_{ij}(s) dB_j(s) \\ &+ \frac{1}{2} \sum_{i, i'=1}^d \int_0^t \frac{\partial^2}{\partial x_i \partial x_{i'}} \Phi(s, X_s) \sum_{j=1}^k \phi_{ij}(s) \phi_{i'j}(s) ds . \end{aligned}$$

Cette formule s'écrit

$$d\Phi(t, X_t) = \frac{\partial}{\partial t} \cdot \Phi dt + \nabla_x \phi \cdot dX_t + \frac{1}{2} \langle dX, D^2 \Phi \cdot dX \rangle ,$$

où

$$\nabla_x \phi \cdot dX_t = \begin{pmatrix} \partial_{x_1} \Phi \\ \vdots \\ \partial_{x_d} \Phi \end{pmatrix} \times (\psi dt + \phi dB) .$$

Nous rappelons que nous avons

$$\langle dB_i, dB_j \rangle = \begin{cases} dt & \text{si } i = j, \\ 0 & \text{sinon} \end{cases}$$

et

$$\langle dt, dt \rangle = 0 .$$

$$\begin{aligned} \langle dX, D^2 \Phi \cdot dX \rangle &= \left\langle \sum_{j'=1}^k \phi_{ij'} dB_{j'}, \sum_{i'=1}^d \frac{\partial^2 \Phi}{\partial x_i \partial x_{i'}} \sum_{j=1}^k \phi_{i'j} dB_j \right\rangle \\ &= \sum_{i, i', j, j'} \frac{\partial^2 \Phi}{\partial x_i \partial x_{i'}} \phi_{ij'} \phi_{i'j} \langle dB_{j'}, dB_j \rangle \\ &= \sum_{i, i', j, j'} \frac{\partial^2 \Phi}{\partial x_i \partial x_{i'}} \phi_{ij'} \phi_{i'j} dt . \end{aligned}$$

$$d\Phi(t, X_t) = \partial_t \Phi \cdot dt + \nabla_x \Phi \cdot dX + \frac{1}{2} \text{tr}(D^2 \Phi \phi \phi^t) dt .$$

8.3.5 Intégration par parties

Soient $B, X, Y \in \mathbb{R}$:

$$\begin{aligned} dX &= \phi dB + \psi dt, \\ dY &= \lambda dB + \mu dt, \end{aligned}$$

avec $\phi, \psi, \lambda, \mu \in M_{\text{loc}}^2$.

$$dXY = X dY + Y dX + \langle dX, dY \rangle .$$

On a donc :

$$\begin{aligned} X(t) Y(t) - X(0) Y(0) &= \int_0^t X(s) [\lambda(s) dB(s) + \mu(s) ds] \\ &\quad + \int_0^t Y(s) [\phi(s) dB(s) + \psi(s) ds] + \int_0^t \phi(s) \lambda(s) ds . \end{aligned}$$

8.4 Formule de Girsanov

8.4.1 Formule de Cameron-Martin

Théorème 8.1 (Cameron-Martin) — Soit $X(t)$, pour $t \geq 0$, une f.a.r.c. gaussienne centrée, et soit $m : \mathbb{R}^+ \rightarrow \mathbb{R}$ une fonction de la forme $m(t) = \mathbb{E}(X(t) Y)$, $t \geq 0$, $Y \in H^X$. Alors

$$\mathbb{E}(F(X + m)) = \mathbb{E}\left(F(X) \cdot \exp\left[Y - \frac{1}{2}\mathbb{E}(Y^2)\right]\right) .$$

8.4.2 Théorème de Girsanov

Soit $\phi \in M_{\text{loc}}^2$. Soit

$$Z(t) = \exp\left[\int_0^t \phi(s) dB(s) - \frac{1}{2}\int_0^t \phi^2(s) ds\right] .$$

$Z(t)$ est une martingale locale.

Théorème 8.2 (Girsanov) — Supposons que $\mathbb{E}(Z(t)) = 1, \forall t$. Alors il existe une proba. \mathbb{Q} définie sur \mathcal{F}_∞ par

$$\mathbb{Q}(A) = \int_A Z(t) d\mathbb{P} \quad \text{pour } A \in \mathcal{F}_t$$

et telle que

$$\bar{B}(t) = B(t) - \int_0^t \phi(s) ds$$

soit, sous \mathbb{Q} , un mouvement brownien.

Lemme 8.1 — Nous avons :

$$\int_0^t \phi^2(s) ds \leq c \quad \Rightarrow \quad \begin{cases} \mathbb{E}(Z(t)) = 1, \\ \mathbb{E}(Z^2(t)) < \infty. \end{cases}$$

Lemme 8.2 (Gronwall) — Soit $t \mapsto x(t)$ telle que

$$x(t) \leq a + b \int_0^t x(s) ds \quad \forall t, a, b > 0.$$

Alors

$$x(t) \leq ae^{bt}.$$

8.4.3 Critères

Proposition 8.8 (Critère de Novikov) — Nous avons :

$$\mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^t \phi^2(s) ds \right) \right] < \infty \quad \Rightarrow \quad \begin{cases} \mathbb{E}(Z(t)) = 1, \\ Z \text{ martingale.} \end{cases}$$

Proposition 8.9 — S'il existe $a, c > 0$ tels que

$$\mathbb{E} \left[\exp(a\phi^2(s)) \right] < c \quad \forall s \leq t,$$

alors

$$\mathbb{E}(Z(t)) = 1.$$

Processus de comptage

9.1 Rappels concernant les martingales

Théorème 9.1 (Formule de décomposition de Doob-Meyer) — Si $X(t)$ est une sous-martingale, alors il existe un processus cadlag¹, prévisible et croissant $\Lambda(t)$ tel que

$$M(t) = X(t) - \Lambda(t)$$

soit une martingale uniformément intégrable.

Remarque — Λ est la somme des espérances conditionnelles (par rapport au passé) des accroissements de X (qui ne peut décroître puisque il est une sous-martingale). M , elle, est la somme des accroissements moins leurs espérances conditionnelles. Cette orthogonalité entre processus prévisibles (à variation finie) et martingales assure à cette décomposition de Doob-Meyer son unicité.

9.2 Processus à variation prévisible

Proposition 9.1 — Si $M(t)$ est une martingale, alors $M^2(t)$ est une sous-martingale.

Proposition 9.2 — Soit $M(t)$ une martingale. Alors

$$M^2(t) = \mathcal{M}_t + \langle M \rangle_t ,$$

avec \mathcal{M}_t martingale et $\langle M \rangle_t$ processus prévisible croissant défini par

$$\langle M \rangle_t = \lim_{|\delta| \rightarrow 0} \sum_i \mathbb{E} [(M_{t_{i+1}} - M_{t_i})^2 | \mathcal{F}_{t_i}]$$

et appelé **processus prévisible croissant** associé à $M(t)$.

1. Continu à droite avec une limite à gauche.

Proposition 9.3 — Soit $M(t)$ une martingale. Alors

$$\mathbb{V} (dM(t) \mid \mathcal{F}_{t-}) = d \langle M \rangle (t) .$$

Proposition 9.4 — Si M_1 et M_2 sont deux martingales (localement) de carré intégrable, il existe un processus prévisible unique (localement) intégrable et à variation bornée, noté $\langle M_1, M_2 \rangle$, tel que $M_1 M_2 - \langle M_1, M_2 \rangle$ soit une martingale (locale), nulle à l'instant 0. $\langle M_1, M_2 \rangle$ est appelé le **processus prévisible de covariation** de M_1 et M_2 .

Nous avons :

$$\text{Cov}(dM_1(t), dM_2(t) \mid \mathcal{F}_{t-}) = d \langle M_1, M_2 \rangle (t) .$$

Proposition 9.5 — Le processus prévisible de covariation est bilinéaire et symétrique, tout comme une covariance ordinaire :

$$\begin{aligned} \langle aM_1 + bM_2, M_3 \rangle &= a \langle M_1, M_3 \rangle + b \langle M_2, M_3 \rangle , \\ \langle M_1, M_2 \rangle &= \langle M_2, M_1 \rangle . \end{aligned}$$

M_1 et M_2 sont dites **orthogonales** ssi $\langle M_1, M_2 \rangle = 0$.

Définition 9.1 — À tout processus cadlag X , on peut associer un **processus de saut** ΔX , défini par

$$\Delta X(t) = X(t) - X(t^-) .$$

Proposition 9.6 — Si M_1 et M_2 sont deux martingales (localement) de carré intégrable telles que

$$\Delta M_1 \cdot \Delta M_2 = 0$$

(i.e. n'ayant aucun temps de saut en commun), alors

$$\langle M_1, M_2 \rangle = 0 .$$

9.3 Processus de comptage

9.3.1 Cas univarié

Définition 9.2 — Un **processus de comptage** N est un processus cadlag, adapté, nul en zéro, croissant et ayant des sauts d'amplitude 1.

Définition 9.3 — Soit $N(t)$ un processus de comptage. C'est (par définition) une sous-martingale locale. Par conséquent, il existe un processus $\Lambda(t)$ prévisible, croissant, continu à droite et nul en zéro tel que

$$M(t) = N(t) - \Lambda(t)$$

soit une martingale.

$\Lambda(t)$ s'appelle le **compensateur** de $N(t)$, ou encore sa **projection prévisible**.

Proposition 9.7 — Soient N un processus ponctuel de dimension 1, et Λ son compensateur. Si N est absolument continu, alors N possède une **intensité** λ , i.e. il existe un processus prévisible λ tel que

$$\Lambda(t) = \int_0^t \lambda(s) \, ds$$

pour tout t . L'intensité est définie par :

$$\lambda(s) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P}(N(s+\epsilon) - N(s) \geq 1 \mid \mathcal{F}_s) .$$

Proposition 9.8 — Soit N un processus de comptage et Λ son compensateur. Le processus prévisible associé à la martingale locale de carré intégrable $M = N - \Lambda$ (ou encore le compensateur de M^2) vaut

$$\begin{aligned} \langle M \rangle &= \Lambda - \int \Delta\Lambda \, d\Lambda \\ &= \int (1 - \Delta\Lambda) \, d\Lambda \end{aligned}$$

et en particulier, si Λ est continu,

$$\langle M \rangle = \Lambda .$$

Théorème 9.2 (Théorème de l'innovation) — Soit N un processus de comptage adapté par rapport à deux filtrations $(\mathcal{F}_t)_t$ et $(\mathcal{G}_t)_t$ telles que $\mathcal{F}_t \subseteq \mathcal{G}_t$. N a pour intensité λ par rapport à $(\mathcal{G}_t)_t$. Alors il existe un processus $\tilde{\lambda}$ prévisible par rapport à $(\mathcal{F}_t)_t$ et tel que :

$$\tilde{\lambda}(t) = \mathbb{E}[\lambda(t) \mid \mathcal{F}_{t-}] .$$

Remarque — $\tilde{\lambda}$ est le processus d'intensité de N par rapport à $(\mathcal{F}_t)_t$.

9.3.2 Cas multivarié

Définition 9.4 — Un processus de comptage r -dimensionnel $N = \{N_i : i = 1, \dots, r\}$ est appelé **processus de comptage multivarié** si chacune de ses composantes est un processus de comptage univarié et s'il ne peut y avoir simultanément des sauts de deux (ou plus) de ses composantes.

Proposition 9.9 — Soit $N = \{N_i : i = 1, \dots, r\}$ un processus de comptage multivarié. Alors :

1. il existe des processus prévisibles Λ_i continus à droite, croissants, nuls à l'instant $t = 0$, tels que les $N_i - \Lambda_i$ soient des martingales localement de carré intégrable ;
2. $N. = \sum_{i=1}^r N_i$ est un processus de comptage de compensateur $\Lambda. = \sum_{i=1}^r \Lambda_i$.

Proposition 9.10 — Soit N un processus de comptage multivarié et Λ son compensateur. Le processus prévisible associé à la martingale locale de carré intégrable $M = N - \Lambda$ (ou encore le compensateur de M^2) vaut

$$\begin{aligned} \langle M_i \rangle &= \Lambda_i - \int \Delta \Lambda_i \, d\Lambda_i \\ &= \int (1 - \Delta \Lambda_i) \, d\Lambda_i , \end{aligned}$$

$$\langle M_i, M_j \rangle = - \int \Delta \Lambda_i \, d\Lambda_j \quad (i \neq j) .$$

En particulier, si Λ est continu,

$$\begin{aligned} \langle M_i \rangle &= \Lambda_i , \\ \langle M_i, M_j \rangle &= 0 \quad (i \neq j) . \end{aligned}$$

9.4 Théorème de la limite centrale

Théorème 9.3 (Théorème de Rebolledo) — Si M_n est une suite de martingales, et si :

- (i) $\langle M_n \rangle_t$ converge en probabilité vers v_t déterministe ;
- (ii) $\forall \epsilon, \exists M_{n,\epsilon}$ suite de martingales telles que $M_n - M_{n,\epsilon}$ n'ait aucune amplitude supérieure à ϵ ,

alors $M_n(t)$ a une limite $M(t)$ de processus croissant v_t , donc $M(t)$ est un processus gaussien :

$$\frac{M_n(t)}{v_t} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) .$$

9.5 Résidus

Proposition 9.11 — Soit le processus martingale

$$M_i(t) = \int_0^t dN_i(s) - \int_0^t \lambda_i(s) \, ds ,$$

et $H_i(t)$ un processus prévisible et localement borné. Alors :

$$R_i(t) = \int_0^t H_i(s) dM_i(s) .$$

est une martingale de moyenne nulle vérifiant

$$\text{Cov}[R_i(s), R_j(t)] = 0$$

pour $i \neq j$, et ceci bien que R_i et R_j ne soient pas indépendants (à moins que H_i et H_j ne le soient).

De plus,

$$\mathbb{V}[R_i(t)] = \mathbb{E} \left[\left(\int_0^t H_i(u) \lambda_i(u) du \right)^t \left(\int_0^t H_i(u) \lambda_i(u) du \right) \right] .$$

9.6 Théorie du produit intégral (ou produit infini)

Théorème 9.4 — Soit $X(s)$ un processus cadlag, nul en 0, et à variation bornée. On obtient une mesure additive en posant

$$X(]s,t]) = X(t) - X(s) .$$

Définition 9.5 — Soit une partition $t_0 = s < t_1 < \dots < t_n = t$. Son pas est

$$|\delta| = \sup_i |t_i - t_{i-1}| .$$

Définition 9.6 — On appelle **produit intégral** (ou **produit infini**)

$$\begin{aligned} \mathcal{P}_s^t(1 + dX) &= \mathcal{P}_{]s,t]}(1 + dX) \\ &= \lim_{|\delta| \rightarrow 0} \prod_{i=1}^n [1 + X(]t_{i-1}, t_i]) \end{aligned}$$

qui est indépendante de la suite des (δ) .

Propriété 9.1 — Pour $s \leq u \leq t$,

$$\mathcal{P}_{]s,t]}(1 + dX) = \mathcal{P}_{]s,u]}(1 + dX) \cdot \mathcal{P}_{]u,t]}(1 + dX) .$$

Propriété 9.2 — $\mathcal{P}_{]s,t]}$ est une fonction de t continue à droite.

Propriété 9.3 — $\mathcal{P}_{]s,s]}(1 + dX) = 1$ et $\mathcal{P}_{]s,t]}(1 + dX) \longrightarrow 1$ ($t \rightarrow s$).

Propriété 9.4 — Si $X(t)$ est continu, alors

$$\mathcal{P}_{]0,t]}(1 + dX) = e^{X(t)} .$$

Théorème 9.5 — On suppose que $\mathcal{P}(1 + dX)$ existe et est une fonction cadlag à variation localement bornée. Alors c'est l'unique solution de l'équation-intégrale

$$Y(t) = 1 + \int_{s \in [0,t]} Y(s-) X(ds) .$$

Théorème 9.6 (Duhamel) — Soient $Y = \mathcal{P}(1 + dX)$ et $Y' = \mathcal{P}(1 + dX')$. Alors

$$Y(t) - Y'(t) = \int_{s \in [0,t]} \mathcal{P}_{]0,s]}(1 + dX) \cdot [X(ds) - X'(ds)] \cdot \mathcal{P}_{(s,t]}(1 + dX') .$$

Si $Y'(t)$ est non singulière, alors

$$\begin{aligned} \frac{Y(t)}{Y'(t)} - 1 &= \int_{s \in [0,t]} \mathcal{P}_{]0,s]}(1 + dX) \cdot [X(ds) - X'(ds)] \cdot \left[\mathcal{P}_{]0,s]}(1 + dX') \right]^{-1} \\ &= \int_0^t \frac{Y(s-)}{Y'(s)} [X(ds) - X'(ds)] . \end{aligned}$$

Théorème 9.7 (Équation de Volterra) — Soient Z et W des fonctions cadlag. À W donné, l'unique solution Z de l'équation de Volterra

$$Z(t) = W(t) + \int_0^t Z(s-) X(ds)$$

est

$$\begin{aligned} Z(t) &= W(t) + \int_0^t W(s-) X(ds) \cdot \mathcal{P}_{(s,t]}(1 + dX) \\ &= W(0) \cdot \mathcal{P}_{]0,t]}(1 + dX) + \int_0^t W(ds) \cdot \mathcal{P}_{(s,t]}(1 + dX) . \end{aligned}$$

9.7 Entr'aperçu d'une approche markovienne des processus de comptage

Proposition 9.12 — Soit $X(t)$, $t \in [0,1]$ un processus de Markov continu à droite et d'espace d'états fini. Soit $N^{(h,j)}(t)$ le nombre de transitions directes de l'état h à l'état j

($h \neq j$) dans l'intervalle $[0, t]$. On suppose que des intensités de transition (de l'état h à l'état j , $h \neq j$) localement intégrables existent : soient $\alpha^{(hj)}(t)$ ces intensités.

Alors le processus d'intensité de N par rapport à $\mathcal{F}_t = \sigma(X(0), N(s), s \leq t)$ est

$$\alpha^{(hj)}(t) Y_h(t)$$

où $Y_h(t) = \mathbf{1}_{\{X(t^-)=h\}}$.

Remarque — Le processus de comptage $N = (N^{(hj)}(\cdot), h \neq j)$ et $X(0)$ sont « équivalents », dans le sens que l'observation de $X(u)$ pour $0 \leq u \leq t$ fournit la même information que l'observation conjointe de $X(0)$ et de $N(u)$ pour $0 \leq u \leq t$.

Quatrième partie

**ÉQUATIONS
DIFFÉRENTIELLES
STOCHASTIQUES**

10

Introduction

10.1 Existence et unicité de solutions fortes

Équation différentielle stochastique (EDS) :

$$\begin{cases} dX = f(X) dt + g(X) dB(t) \\ X_0 \text{ condition initiale} \end{cases}$$

Ceci s'écrit encore

$$X(t) = X_0 + \int_0^t f(X(s)) ds + \int_0^t g(X(s)) dB(s).$$

$f(t, x)$ est appelée **dérive (drift)** de l'EDS, et $g(t, x)$ **coefficient de diffusion** de l'EDS.

Théorème 10.1 — Soient $(\Omega, \mathcal{A}, \mathbb{P})$, $(B_t, t \geq 0)$ un $(\mathcal{F}_t)_t$ -brownien sur Ω , X_0 indépendant de $(B_t, t \geq 0)$. On suppose que

$$|f(t, x) - f(t, y)| + |g(t, x) - g(t, y)| \leq K|x - y| \quad \forall t, x, y.$$

Alors il existe une unique solution X de l'EDS — et $X \in M^2$. C'est une solution forte au sens où X est une fonction mesurable de X_0 et de B .

10.2 Exemples

Ornstein-Uhlenbeck

$$\begin{aligned} dV_t &= -\alpha V_t dt + dB_t, \\ dX &= \alpha X dt + \sigma dB \end{aligned}$$

dont la solution est

$$X(t) = X_0 \exp \left[\sigma B(t) + \left(\alpha - \frac{\sigma^2}{2} \right) t \right].$$

EDS

$$\begin{cases} dX_t = \sqrt{1 + X_t^2} dB_t + (\sqrt{1 + X_t^2} + \frac{1}{2}X_t) dt \\ X(0) = X_0 \end{cases}$$

$$(\text{sh}y)' = \text{ch}y = \sqrt{1 + \text{sh}^2y}$$

$$Y_t = \text{sh}(B_t)$$

$$\begin{aligned} \text{Ito} : dY_t &= \text{ch}(B_t) + \frac{1}{2}\text{sh}(B_t) dt \\ &= \sqrt{1 + Y_t^2} dB_t + \frac{1}{2}Y_t dt \end{aligned}$$

$$X_t = \text{sh}(B_t + t + \text{Argsh}X_0)$$

Vérification :

$$\begin{aligned} dX_t &= \text{ch}(B_t + t + \text{Argsh}X_0)[dB_t + dt] + \frac{1}{2}\text{sh}(B_t + t + \text{Argsh}X_0) dt \\ &= \sqrt{1 + X_t^2}(dB_t + dt) + \frac{1}{2}X_t dt . \end{aligned}$$

Autre EDS Soit

$$dx = x^2 dt .$$

Une solution est

$$x(t) = \frac{x_0}{1 - x_0 t} .$$

Soit

$$X(t) = \frac{1}{1 - B(t)} .$$

$$\text{Ito} : dX_t = \left(\frac{1}{1 - B_t}\right)^2 dB_t + \frac{2}{2}\left(\frac{1}{1 - B_t}\right)^3 dt$$

$$dX_t = X_t^2 dB_t + X_t^3 dt .$$

10.3 Solutions faibles d'EDS

Une **solution faible** d'EDS est un triplet $((X, B), (\Omega, \mathcal{F}, \mathbb{P}), (\mathcal{F}_t)_t)$ tel que $\forall t$,

$$X_t = X_0 + \int_0^t f(X_s) ds + \int_0^t g(X_s) dB_s \quad \text{p.s.}$$

Proposition 10.1 — Soit b borélienne de $\mathbb{R}^+ \times \mathbb{R}^d$ dans \mathbb{R}^d telle que $|b(t, x)| \leq K(1 + |x|)$. Soit μ une proba sur \mathbb{R}^d . Soit l'EDS $dX_t = b(t, X_t) dt + dB_t$. Alors l'EDS a une solution faible X de loi initiale μ .

10.3.1 Modèle de Cox – Ingersoll – Ross

$$\begin{cases} dX_t = c(\theta - X_t) dt + \sigma\sqrt{X_t} dB_t \\ X(0) = x_0 > 0 \end{cases}$$

avec c, θ, σ constantes.

Dans le cas où $4c\theta = \sigma^2$, alors $X = Y^2$ avec

$$\begin{cases} dY_t = \frac{\sigma}{2} dB_t - \frac{c}{2} Y_t dt \\ Y(0) = \sqrt{x_0} \end{cases}$$

qui est un processus d'Ornstein-Uhlenbeck. En effet, d'après Ito,

$$\begin{aligned} dX_t &= 2Y_t dY_t + \frac{1}{2} 2 \langle dY_t, dY_t \rangle \\ &= 2Y_t \left(\frac{\sigma}{2} dB_t - \frac{c}{2} Y_t dt \right) + \frac{\sigma^2}{4} dt \\ &= \sigma\sqrt{X_t} + \underbrace{\left(\frac{\sigma^2}{4} - cX_t \right)}_{c(\theta - X_t)} dt. \end{aligned}$$

10.3.2 Absolue continuité de la loi de diffusions sous changement de dérive

On considère 2 EDS :

$$\begin{cases} dX_t = \sigma(X_t) dB_t + b_1(X_t) dt & \text{(EDS 1)} \\ dY_t = \sigma(Y_t) dB_t + b_0(Y_t) dt & \text{(EDS 0)} \\ X(0) = Y(0). \end{cases}$$

Soit $\tilde{\mathbb{P}}_1 = X \circ \mathbb{P}$ la loi de X sous \mathbb{P} , et $\tilde{\mathbb{P}}_0 = Y \circ \mathbb{P}$ la loi de Y sous \mathbb{P} . On va montrer que $\tilde{\mathbb{P}}_1 \ll \tilde{\mathbb{P}}_0$ (et même équivalentes) et calculer

$$\frac{d\tilde{\mathbb{P}}_1}{d\tilde{\mathbb{P}}_0}(y),$$

où $y = (y_t, t \leq T) \in \mathcal{C}([0, T], \mathbb{R})$.

Soit

$$Z(T) = \exp \left[\int_0^T h(Y_s) dB_s - \frac{1}{2} \int_0^T h^2(Y_s) ds \right]$$

avec

$$h(y) = \frac{b_1 - b_0}{\sigma} y.$$

On définit \mathbb{Q} par $d\mathbb{Q} = Z d\mathbb{P}$ sur \mathcal{F}_t (Girsanov). On fait les deux hypothèses suivantes :

- $\sigma > 0$;
- $\mathbb{E}(Z(T)) = 1$ (i.e. vrai si martingale).

Sous \mathbb{Q} ,

$$\bar{B}(t) = B(t) - \int_0^t h(Y_s) ds \quad \text{est un brownien}$$

$$\begin{aligned} dY_t &= \sigma(Y_t) dB_t + b_0(Y_t) dt \\ &= \sigma(Y_t) d\bar{B}_t + b_1(Y_t) dt . \end{aligned}$$

On fait l'hypothèse supplémentaire qu'il existe une unique solution en loi de (EDS 1).
Alors $X \circ \mathbb{P} = Y \circ \mathbb{Q}$.

D'autre part, Y est sous \mathbb{Q} solution de (EDS 1).

Soit $\psi : \mathcal{C}([0, T], \mathbb{R}) \rightarrow \mathbb{R}$ continue bornée.

$$\begin{aligned} \int_{\mathcal{C}} \psi d\tilde{\mathbb{P}}_1 &= \mathbb{E}^{\mathbb{P}}(\psi(X)) \\ &= \mathbb{E}^{\mathbb{Q}}(\psi(Y)) \\ &= \mathbb{E}^{\mathbb{P}}(\psi(Y)Z) \\ &= \mathbb{E}^{\mathbb{P}}[\psi(Y)\mathbb{E}^{\mathbb{P}}(Z | Y)] . \end{aligned}$$

D'où

$$\begin{aligned} Z(t) &= \exp \left[\int_0^T h(Y_s) dB_s - \frac{1}{2} \int_0^T h^2(Y_s) ds \right] \\ &= \exp \left[\int_0^T h(Y_s) \frac{dY_s - b_0(Y_s) ds}{\sigma(Y_s)} - \frac{1}{2} \int_0^T h^2(Y_s) ds \right] \\ &= \exp \left[\int_0^T \frac{h(Y_s)}{\sigma(Y_s)} dY_s - \int_0^T \left[\frac{h(Y_s)b_0(Y_s)}{\sigma(Y_s)} + \frac{1}{2}h^2(Y_s) \right] ds \right] \end{aligned}$$

...et

$$\begin{aligned} &\frac{h(Y_s) b_0(Y_s)}{\sigma(Y_s)} + \frac{1}{2}h^2(Y_s) \\ &= \frac{b_1(Y_s) - b_0(Y_s)}{\sigma(Y_s)} \left(\frac{b_0(Y_s)}{2} + \frac{b_1(y_s) - b_0(Y_s)}{2\sigma(Y_s)} \right) \\ &= \frac{b_1^2(Y_s) - b_0^2(Y_s)}{2\sigma^2} . \end{aligned}$$

D'où

$$\mathbb{E}^{\mathbb{P}}[Z | Y] = \exp \left[\int_0^T \frac{b_1(Y_s) - b_0(Y_s)}{\sigma^2(Y_s)} dY_s - \frac{1}{2} \int_0^T \frac{b_1^2(Y_s) - b_0^2(Y_s)}{\sigma^2(Y_s)} ds \right] .$$

Remarque — Si on a des coefficients de diffusion différents, i.e.

$$\begin{cases} dX_t = \sigma(X_t) dB_t + b(X_t) dt , \\ dY_t = \tau(Y_t) dB_t + b(Y_t) dt , \\ X(0) = Y(0) , \end{cases}$$

alors on perd l'absolue continuité — les probabilités deviennent même étrangères.

10.3.3 Équations linéaires

Le premier exemple est le suivant :

$$(EDSL) \quad \begin{cases} dX_t = (A_t X_t + a_t) dt + \sigma_t dB_t \\ X(0) = \zeta, \end{cases}$$

avec $X, a \in \mathbb{R}^d$ — $A, \sigma \in \mathcal{M}_{d \times d}$ et $B \in \mathbb{R}^k$.

Le deuxième exemple est l'équation différentielle ordinaire linéaire (EDOL) :

$$d\zeta_t = (A_t \zeta_t + a_t) dt.$$

Une solution fondamentale de l'EDOL est $\phi \in \mathcal{M}_{d \times d}$ telle que

$$\begin{cases} \dot{\phi}_t = A_t \phi_t \\ \phi_0 = \text{Id} \end{cases}$$

On a alors

$$\zeta_t = \phi_t \left[\zeta + \int_0^t \phi_s^{-1} a_s ds \right].$$

La solution de l'EDSL est

$$X_t = \phi(t) \left[X_0 + \int_0^t \phi^{-1}(s) a(s) ds + \int_0^t \phi^{-1}(s) \sigma(s) dB_s \right].$$

D'après Ito :

$$\begin{aligned} dX_t &= \dot{\phi}(t) \left[X_0 + \int_0^t \phi^{-1}(s) a(s) ds + \int_0^t \phi^{-1}(s) \sigma(s) dB_s \right] dt \\ &\quad + \phi(t) d \left[X_0 + \int_0^t \phi^{-1}(s) a(s) ds + \int_0^t \phi^{-1}(s) \sigma(s) dB_s \right] \\ &= A_t \phi_t \left[X_0 + \int_0^t \phi^{-1}(s) a(s) ds + \int_0^t \phi^{-1}(s) \sigma(s) dB_s \right] \\ &\quad + \phi_t [\phi_t^{-1} a_t dt + \phi_t^{-1} \sigma_t dB_t] \\ &= A_t X_t dt + a_t dt + \sigma_t dB_t \\ &= \text{EDSL}. \end{aligned}$$

10.3.4 Autre EDS

$$\begin{cases} dX_t = A_t X_t dt + \sigma_t X_t dB_t, \\ X(0). \end{cases}$$

La solution est

$$X_t = X_0 \exp \left[\int_0^t \sigma_s dB_s + \int_0^t \left(A_s - \frac{\sigma_s^2}{2} \right) ds \right].$$

10.3.5 Pont brownien

Le brownien est conditionné pour revenir à l'origine à l'instant 1. Notons $B(t) = [B(t) - tB(1)] + tB(1)$. On a

$$\begin{aligned} \text{cov} (B(1), B(t) - tB(1)) &= \mathbb{E} (B(1) \cdot B(t)) - \mathbb{E} (t B^2(1)) \\ &= t - t \times 1 \\ &= 0 . \end{aligned}$$

D'où $B(t)$ est une somme de 2 processus indépendants.

Définition 10.1 — *Le processus $(B(t) - tB(1))_{0 \leq t \leq 1}$ est appelé **pont brownien**.*

Il est indépendant de $B(1)$ et est noté $(B_1^{0 \rightarrow 0}(t))_{0 \leq t \leq 1}$. C'est un processus gaussien centré de covariance

$$\rho(s, t) = s \wedge t - st .$$

Illustration — On va regarder $B_T^{a \rightarrow b}(t)$, *i.e.* le brownien qui part de a et parvient en b au temps T . Soit

$$\begin{cases} dX_t = \frac{b-X_t}{T-t} dt + dB_t , \\ X(0) = a . \end{cases}$$

C'est une EDSL ; posons $\phi_t = 1 - t/T$. La solution est

$$X_t = a\left(1 - \frac{t}{T}\right) + b\frac{t}{T} + \underbrace{(T-t) \int_0^t \frac{dB_s}{T-s}}_{B_T^{a \rightarrow b}(t)} .$$

$B_T^{a \rightarrow b}(t)$ est un processus gaussien centré de covariance

$$\rho(s, t) = s \wedge t - \frac{st}{T} .$$

Propriétés des EDS

11.1 Caractère markovien des équations différentielles stochastiques

11.1.1 Propriété de Markov des solutions des EDS

Introduction Soient $(\Omega, \mathcal{A}, \mathbb{P}), (\mathcal{F}_t, t \geq 0)$ une filtration, $X_t, t \geq 0$ un processus adapté dans \mathbb{R}^d .

Définition 11.1 — *Le processus X est un **processus de Markov** par rapport à $(\mathcal{F}_t, t \geq 0)$ si $\forall t \geq 0, \forall h > 0, \forall A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mathbb{P}(X(t+h) \in A \mid \mathcal{F}_t) = \mathbb{P}(X(t+h) \in A \mid X(t)) \quad p.s.$$

c.-à-d. que la loi du futur, sachant le passé, ne dépend que du présent.

Définition 11.2 — *Le processus X est un **processus de Markov homogène** si la loi conditionnelle ne dépend que de h (et pas de t). On note $Q(h, x; dy)$ la loi conditionnelle :*

$$\mathbb{P}(X(t+h) \in A \mid X(t) = x) = Q(h, x; A).$$

$Q : \mathbb{R}^+ \times \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0,1]^{\otimes d}$ telle que :

- (i) $\forall A, (h, x) \mapsto Q(h, x; A)$ est mesurable;
- (ii) $\forall h, x, A \mapsto Q(h, x; A)$ est une proba sur \mathbb{R}^d .

$$Q(h, x; A) = \int_{y \in A} Q(h, x; dy)$$

$Q(h, x; dy)$ est une **probabilité de transition**. Soit $0 = t_0 < t_1 < \dots < t_k$. La loi de $X(t_1), X(t_2), \dots, X(t_k)$ partant de $X(0) = x$ est

$$Q(t_1, x; dx_1) \times Q(t_2 - t_1, x_1; dx_2) \times \dots \times Q(t_k - t_{k-1}, x_{k-1}; dx_k).$$

Proposition 11.1 (Équation de Chapman-Kolmogorov) —

$$Q(s+t, x; A) = \int_{y \in \mathbb{R}^d} Q(s, x; dy) \cdot Q(t, y; A).$$

En effet,

$$\begin{aligned} Q(s+t, x; A) &= \mathbb{P}(X(s+t) \in A \mid X(0) = x) \\ &= \mathbb{E} \left[\mathbb{P}(X(s+t) \in A \mid \mathcal{F}_s) \mid X(0) = x \right] \\ &= \mathbb{E} \left[\mathbb{P}(X(s+t) \in A \mid X_s) \mid X(0) = x \right] \\ &= \int_{y \in \mathbb{R}^d} Q(t, y; A) \cdot Q(s, x; dy). \end{aligned}$$

Proposition 11.2 — *Un \mathcal{F} -mouvement brownien est un \mathcal{F} -processus de Markov.*

Proposition 11.3 — *Soit T un \mathcal{F} -temps d'arrêt.*

$$W(t) = B(t+T) - B(T), \quad t \geq 0$$

est un mouvement brownien.

Équations progressive et rétrograde de Kolmogorov Soit

$$(EDS) \quad \begin{cases} dX_t = f(X_t) dt + g(X_t) dB_t \\ X_0 \in \mathcal{F}_0, \end{cases}$$

avec f et g globalement lipschitziennes, $X \in \mathbb{R}^d$, $B \in \mathbb{R}^k$, $g = (g_{ij})_{i \leq d, l \leq k}$. L'unique solution X de (EDS) est un \mathcal{F} -processus de Markov.

Soit l'EDS

$$d\phi(X_t) = [\phi'(X_t)f(X_t) + \frac{1}{2}\phi''(X_t)g^2(X_t)] dt + \phi'(X_t)g(X_t) dB_t.$$

Soit

$$L\phi(x) = \phi'(x) f(x) + \frac{1}{2}\phi''(x) g(x)^2.$$

L est l'opérateur différentiel linéaire du second ordre; c'est le **générateur infinitésimal** de $(X_t)_t$. Dans la cas (général) vectoriel :

$$\phi(X_t) = \phi(X_0) + \int_0^t L\phi(X_s) ds + \int_0^t \sum_{l=1}^k \phi(X_s) dB_l(s),$$

avec

$$L\phi(x) = \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 \phi}{\partial x_i \partial x_j} + \sum_{i=1}^d f_i(x) \frac{\partial}{\partial x_i} \phi(x)$$

où

$$gg^* = a ,$$

$$M_l = \sum_{i=1}^d g_{il}(x) \frac{\partial}{\partial x_i} .$$

$\phi(X_t) - \int_0^t L\phi(X_s) ds$ est :

- une martingale locale si $\phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$;
- une martingale si $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R})$.

$$\begin{aligned} \mathbb{E}_x(\phi(X_t)) &= \phi(x) + \int_0^t \mathbb{E}_x(L\phi(X_s) ds) \\ &= Q(t, x; \phi) \end{aligned}$$

$$\int_0^t \mathbb{E}_x(L\phi(X_s)) ds = \int_0^t Q(s, x; L\phi) ds$$

car

$$\begin{aligned} Q(t, x; A) &= \int_{y \in A} Q(t, x; dy) \\ \int \phi(y) Q(t, x; dy) &= Q(t, x; \phi) . \end{aligned}$$

$$\begin{aligned} \partial_t Q(t, x; \phi) &= Q(t, x; L\phi) \\ \frac{\partial}{\partial t} Q(t, x; dy) &= L_y^* Q(t, x; dy) \end{aligned}$$

où L_y est l'opérateur L pour la variable y et L^* l'adjoint de L (transposition).

$$Q(h, x; L\phi) = \int Q(t, x; dy) (L\phi)(y) .$$

$$L^*\phi(x) = \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (a_{ij}(x)\phi(x)) - \sum_{i=1}^d \frac{\partial}{\partial x_i} [f_i(x)\phi(x)] .$$

Prendre l'adjoint est une opération linéaire : si

$$K\phi(x) = f_i(x) \frac{\partial}{\partial x_i} \phi(x) ,$$

alors $K^*\psi$ est défini par :

$$\begin{aligned} \int \psi \cdot K\phi &= \int K^*\psi \cdot \phi \\ &\parallel \\ \int f_i(x) \frac{\partial}{\partial x_i} \phi(x) \cdot \psi(x) dx &= - \int \phi(x) \frac{\partial}{\partial x_i} (f_i(x)\psi(x)) dx . \end{aligned}$$

D'où

$$K^* \psi = -\frac{\partial}{\partial x_i} (f_i(x) \psi(x)) .$$

On a donc l'équation **progressive (forward) de Kolmogorov** :

$$\frac{\partial}{\partial t} Q(t, x; dy) = L_y^* Q(t, x; dy) .$$

D'autre part, $Q(t, x, dy) \rightarrow \delta_x$ quand $t \searrow 0$.

$$\begin{aligned} \partial_t \left[\int Q(t, x; dy) Q(s, y; dz) \right] &= \partial_t [Q(s+t, x; dz)] \\ \Leftrightarrow \int L_y^* Q(t, x; dy) Q(s, y; dz) &= \partial_t [Q(s+t, x; dz)] \\ \Leftrightarrow \int Q(t, x; dy) L_y Q(s, y; dz) &= \partial_t [Q(s+t, x; dz)] . \end{aligned}$$

On a donc l'équation **rétrograde (backward) de Kolmogorov** (pour $t = 0$) :

$$\frac{\partial}{\partial t} Q(t, x; dz) = L_x Q(t, x; dz) .$$

11.1.2 Générateurs et EDS

$$X_t = X_0 + \int_0^t \sigma(s, X_s) dB_s + \int_0^t b(s, X_s) ds . \quad (11.1)$$

$$d\phi(t, X_t) = \left(\frac{\partial \phi}{\partial t} + L_t \phi \right) (t, X_t) dt + \sum_l \left[\left(\sum_i \frac{\partial \phi}{\partial x_i} \sigma_{il} \right) dB_l \right] .$$

$$L_t \phi(x) = \frac{1}{2} \sum_{i,j} a_{ij}(t, x) \frac{\partial^2}{\partial x_i \partial x_j} \phi(x) + \sum_{i=1}^d b_i(t, x) \frac{\partial}{\partial x_i} \phi(x) ,$$

où $a = \sigma \sigma^*$.

Problème de Cauchy

$$\text{(Cauchy)} \quad \begin{cases} \frac{\partial}{\partial t} v = -L_t v - kv + g \\ v(T, x) = f(x) \quad (\text{condition finale}), \end{cases}$$

avec $v = v(t, x)$, $t \in [0, T]$, $x \in \mathbb{R}^d$, $k = k(t, x)$, $g = g(t, x)$, $k \geq 0$.

On fait les hypothèses suivantes :

- b, σ continues sur $[0, T] \times \mathbb{R}^d$ et sous-linéaires¹ ;
- l'EDS (11.1) a une unique solution faible ;
- $f(x), g(t, x)$ et $k(t, x)$ sont continues, f et g à croissance sous-polynômiale, *i.e.*

$$|f(x)| + |g(t, x)| \leq K(1 + |x|^\lambda) ;$$

- a, b et k sont bornées ;
- L_t est uniformément elliptique : $\exists \delta > 0, \forall t, x,$

$$\sum a_{ij}(t, x) \zeta_i \zeta_j \geq \delta |\zeta|^2 .$$

La solution est la **formule de Feynman-Kac**

$$v(t, x) = \mathbb{E}_{t, x} \left\{ f(X_T) \exp \left[- \int_t^T k(u, X_u) du \right] + \int_t^T \left[g(s, X_s) \exp \left(- \int_t^s k(u, X_u) du \right) \right] ds \right\} .$$

Le problème de Cauchy a une unique solution v telle que $|v(t, x)| \leq c(1 + |x|^\mu)$, qui est donnée par la représentation de Feynman-Kac.

Problème de Dirichlet Soient D un ouvert de \mathbb{R}^d , b, σ indépendants de t . On cherche $u \in C(\bar{D})$ solution de

$$\begin{cases} Lu - ku = -g & \text{sur } D \\ u|_{\partial D} = f \end{cases}$$

où $f : \partial D \rightarrow \mathbb{R}, g : \bar{D} \rightarrow \mathbb{R}$ et $k : \bar{D} \rightarrow \mathbb{R}^+$.

On fait les hypothèses suivantes :

- les trois premières parmi celles du problème précédent ;
- a, b, k et g sont hölderiennes ;
- L uniformément elliptique.

Le problème de Dirichlet a une unique solution :

$$u(x) = \mathbb{E}_x \left\{ f(X_\tau) \exp \left[- \int_0^\tau k(X_s) ds \right] + \int_0^\tau \left[g(X_t) \exp \left(- \int_0^t k(X_s) ds \right) \right] dt \right\} ,$$

avec $\tau = \inf\{t \geq 0 : X_t \in D^c\}$.

Atteignabilité de points Soit

$$\begin{cases} dX_t = c(\theta - X_t) dt + \sigma \sqrt{X_t} dB_t , \\ X_0 = x > 0 . \end{cases}$$

Proposition 11.4 — Si $2c\theta > \sigma^2$, alors X_t n'atteint pas 0.

1. Ceci signifie que $|b(t, x)| < K(1 + |x|)$.

11.2 Équations différentielles stochastiques rétrogrades

Notations — Nous notons :

- $|\dots|$ la norme dans \mathbb{R} ;
- $\|\dots\|$ la norme dans $\mathbb{R}^{d \times n}$.

Les données sont les suivantes :

- $B = (B_t)_{t \geq 0} = (B_t^1, \dots, B_t^d)$ mouvement brownien d -dimensionnel défini sur $(\Omega, \mathcal{F}, \mathbb{P})$;
- $\mathcal{F}_t = \sigma(B_s, s \geq t)$;
- T temps terminal ;
- la condition finale $\zeta \in \mathcal{L}^2(\Omega, \mathcal{F}_T)$ est à valeurs dans \mathbb{R}^k ;
- la dérive $f(\omega, t, y, z)$ de $\Omega \times [0, T] \times \mathbb{R}^k \times \mathbb{R}^{k \times d}$ est dans \mathbb{R}^k .

On fait l'hypothèse que f est lipschitzienne : $f \in M^2(0, T)$ et $\exists K, \forall y, y', z, z', t,$

$$|f(t, y, z) - f(t, y', z') - f(t, y', z')| \leq K[|y - y'| + \|z - z'\|] .$$

Définition 11.3 — Une *équation différentielle stochastique rétrograde (EDSR)* est de la forme

$$Y_t = \zeta + \int_t^T f(s, Y_s, Z_s) ds - \int_t^T Z_s dB_s .$$

Définition 11.4 — Une *solution d'une EDSR* (ζ, f) est un couple (Y, Z) de processus progressivement mesurables à valeurs dans $\mathbb{R}^k \times \mathbb{R}^{k \times d}$ et vérifiant :

$$\mathbb{E} \left[\int_0^T \|Z_s\|^2 ds \right] < \infty$$

et

$$Y_t = \zeta_t + \int_t^T f(s, Y_s, Z_s) ds - \int_t^T Z_s dB_s . \tag{11.2}$$

On a donc

$$\begin{cases} dY_t = -f(t, Y_t, Z_t) dt + Z_t dB_t , \\ Y_T = \zeta , \end{cases}$$

et donc

$$Y_t = Y_0 - \int_0^t f(s, Y_s, Z_s) ds + \int_0^t Z_s dB_s . \tag{11.3}$$

Définition 11.5 — L'équation (11.2) s'appelle l'équation différentielle stochastique rétrograde (backward), tandis que (11.3) est l'équation différentielle stochastique progressive (forward).

De plus,

- 1) Y_0 est déterministe ;
- 2) Y_t est \mathcal{F}_t -adapté ;
- 3) on a :

$$\int_0^T Z_s dB_s = \zeta + \int_0^T f(s, Y_s, Z_s) ds - \mathbb{E} \left[\zeta + \int_0^T f(s, Y_s, Z_s) ds \right].$$

Proposition 11.5 (Majoration a priori fondamentale) — Soit (Y, Z) solution de l'EDSR(ζ, f). Alors il existe une constante positive c (ne dépendant que de T et k) telle que

$$\mathbb{E} \left[\sup_{[0, T]} |Y_t|^2 + \int_0^T \|Z_t\|^2 dt \right] \leq c \left(\mathbb{E} (|\zeta|^2) + \mathbb{E} \left(\int_0^T f^2(t, 0, 0) dt \right) \right). \quad (11.4)$$

Proposition 11.6 — Soit

$$M_t = \int_0^t u_s dB_s, \quad t \in [0, T]$$

une martingale locale telle que

$$\int_0^T u_s^2 ds < \infty \quad p.s.$$

On note $M_t^* = \sup_{[0, T]} |M_s|$. Alors, $\forall p > 0$, il existe une constante $c_p > 0$ telle que

$$\mathbb{E} \left[(M_t^*)^p \right] \leq c_p \mathbb{E} \left[\left(\int_0^t u_s^2 ds \right)^{\frac{p}{2}} \right].$$

Théorème 11.1 — Sous les mêmes hypothèses que précédemment, il existe une unique solution à l'EDSR(ζ, f) vérifiant (11.4).

Proposition 11.7 — (Y, Z) est solution de l'EDSR(ζ, f) $\Leftrightarrow (Y, Z) = \phi(Y, Z)$, où ϕ est l'application de $\mathcal{B}^2 = (M^2([0, T]))^k \times (M^2([0, T]))^{k \times d}$ dans lui-même, qui à (U, V) associe $\phi(U, V) = (Y, Z)$ — il s'agit d'un théorème du point fixe.

Théorème 11.2 — Il existe une norme sur \mathcal{B}^2 (hilbertienne) telle que ϕ soit une contraction stricte : $\forall \gamma > 0$,

$$\| \! \| \! \| (Y, Z) \! \! \! \|_{\gamma} = \left\{ \mathbb{E} \left[\int_0^T e^{\gamma t} (|Y_t|^2 + \|Z_t\|^2) dt \right] \right\}^{\frac{1}{2}}.$$

Théorème 11.3 (Comparaison) — Soient $k = d = 1$. Soient (ζ, f) et (ζ', f') vérifiant les hypothèses originelles. Supposons que $\zeta \leq \zeta'$ p.s., et que $\forall Y, Z \in \mathbb{R}$,

$$f(t, y, z) \leq f'(t, y, z) \quad dt \otimes d\mathbb{P} \quad p.s.$$

Soient (Y, Z) solution de l'EDSR (ζ, f) et (Y', Z') solution de (ζ', f') . Alors

$$Y_t \leq Y'_t \quad t \in [0, T], \quad p.s.$$

Proposition 11.8 — Soit (Y, Z) solution de l'EDSR (ζ, f) sous les hypothèses originelles. Supposons qu'il existe un temps d'arrêt $\tau \leq T$ tel que :

- a) ζ soit \mathcal{F}_τ mesurable ;
- b) $f(t, y, z) = 0$ sur $[\tau, T]$.

Alors

$$Y_t = Y_{t \wedge \tau}$$

et

$$Z_t = 0 \quad \text{sur } [\tau, T].$$

Proposition 11.9 — Soient, sous les hypothèses originelles, (Y, Z) solution de l'EDSR (ζ, f) et (Y', Z') solution de (ζ', f') . Alors

$$\mathbb{E} \left[\sup_{[0, T]} |Y_t - Y'_t|^2 + \int_0^T \|Z_t - Z'_t\|^2 dt \right] \leq c \left(\mathbb{E} (|\zeta - \zeta'|^2) + \mathbb{E} \left(\int_0^T |f(t, Y_t, Z_t) - f'(t, Y_t, Z_t)|^2 dt \right) \right).$$

11.3 Lien avec les équations aux dérivées partielles semi-linéaires

11.3.1 Rappel sur la formule de feynman-Kac

Soit $\zeta = g(X_T)$, X diffusion construite sur B . Soient :

- $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$;
- $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$;
- b et σ sont supposées continues en $(t, x) \in [0, T] \times \mathbb{R}^d$;
- K tel que $\forall t, x, y$,

$$|b(t, y) - b(t, x)| + \|\sigma(t, y) - \sigma(t, x)\| \leq K|y - x|.$$

On considère l'EDS associée :

$$\text{EDS } \left\{ \begin{array}{l} dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t \end{array} \right.$$

Soit $X_s^{t,x}$ la solution de (EDS) partant de x à l'instant t :

$$X_s^{t,x} = x + \int_t^s b(u, X_u) du + \int_t^s \sigma(u, X_u) dB_u .$$

Propriété 11.1 — (i) Il y a existence et unicité de $X_s^{t,x}$.

(ii) $X_s^{t,x}$ est $\mathcal{F}_s^t = \sigma(B_u - B_t, t \leq u \leq s)$ -mesurable.

(iii) $X_s^{t,x} = X_s^{u, X_u^{t,x}}$, $t \leq u \leq s \leq T$.

À X on associe un générateur différentiel

$$L_t = \frac{1}{2} \sum_{i,j} (\sigma \sigma^t)_{ij}(t, x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b^i(t, x) \frac{\partial}{\partial x_i}$$

et

$$P_{t,s} f(x) = \mathbb{E} [f(X_s^{t,x})] .$$

Proposition 11.10 —

$$\frac{1}{s} [P_{t,s} f(x) - f(x)] \xrightarrow{s \rightarrow t} L_t f(x) .$$

Problème de Cauchy

$$\text{(Cauchy)} \quad \left\{ \begin{array}{l} \frac{\partial}{\partial t} u(t, x) + a(t, x) u(t, x) = r(t, x) u(t, x) \quad t \in [0, T], x \in \mathbb{R}^d \\ u(T, x) = \phi(x) . \end{array} \right.$$

On a :

- $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ continue ;
- $a, r : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ continues.

On cherche une solution dans $C^{1,2}([0, T] \times \mathbb{R}^d)$ vérifiant (Cauchy).

On suppose qu'un tel u existe et vérifie l'hypothèse

$$\text{(H)} \quad \left| \frac{\partial u}{\partial t}(t, x) \right| \leq K_T (1 + |x|^m) \quad m \geq 1 .$$

On suppose que u vérifie (Cauchy) avec l'hypothèse (H). Alors

$$u(t, x) = \mathbb{E} \left\{ \phi(X_T^{t,x}) \exp \left[- \int_t^T r(s, X_s^{t,x}) ds \right] + \int_t^T \left[a(s, X_s^{t,x}) \exp \left(- \int_t^s r(u, X_u^{t,x}) du \right) \right] ds \right\} .$$

Autre résolution Soient :

- $x \in \mathbb{R}^d$;
- $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$;
- $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ continue par rapport à t et x .

On suppose que

$$|b(t, x) - b(t, y)| + \|\sigma(t, x) - \sigma(t, y)\| \leq K \cdot |x - y| .$$

Soit $(X_s^{t,x})_{s \in [t, T]}$ la solution partant de x à l'instant t de l'EDS (\star) :

$$(\star) \quad \begin{cases} dX_s^{t,x} = b(s, X_s^{t,x}) ds + \sigma(s, X_s^{t,x}) dB_s \\ X_t^{t,x} = x \end{cases}$$

Soient :

- $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ continue et vérifiant

$$|g(x)| \leq K(1 + |x|^p) \quad p \geq \frac{1}{2} ;$$

- $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}^{k \times d} \rightarrow \mathbb{R}^k$ **déterministe** et vérifiant

$$|f(s, x, y, z) - f(s, x, y', z')| \leq K(|y - y'| + \|z - z'\|)$$

et

$$|f(t, x, y, z)| \leq K(1 + |x|^p + |y| + \|z\|) .$$

Soit l'EDSR

$$(\star\star) \quad \begin{cases} \zeta = g(X_T^{t,x}) \\ \tilde{f}(\omega, u, y, z) = f(u, X_u^{t,x}, y, z) \end{cases}$$

i.e

$$Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x}) du - \int_s^T Z_u^{t,x} dB_u .$$

Proposition 11.11 — *L'EDSR $(\star\star)$ admet une unique solution $(Y_u^{t,x}, Z_u^{t,x})$, $u \in [0, T]$ pour tout t dans $[0, T]$.*

Proposition 11.12 — *Soit $(Y^{t,x}, Z^{t,x})$ solution de $(\star\star)$. Alors :*

- (1) $Y_s^{t,x}$ est $\mathcal{F}_s^t = \sigma(B_r - B_t, t \leq r \leq s)$ -mesurable ;
- (2) $Y_t^{t,x}$ est déterministe ;
- (3) $\forall h \geq 0, Y_{t+h}^{t,x} = Y_{t+h}^{t+h, X_{t+h}^{t,x}}$;
- (4) Soit $u(t, x) = Y_t^{t,x}$: c'est une fonction continue de (t, x) sur $[0, T] \times \mathbb{R}^d$.

Théorème 11.4 (Kolmogorov) — *S'il existe $q \geq 1$ et $\gamma > 1 + 1 + d$ (i.e. supérieur à la dimension des paramètres) tels que*

$$\mathbb{E} [|X_s^{t,x} - X_{s'}^{t',x'}|^q] \leq K [|t - t'| + |s - s'| + \|x - x'\|_d^2]^\gamma,$$

alors il existe une version continue de $(t, s, x) \mapsto X_s^{t,x}$. Pour tout $p \geq 1$,

$$\mathbb{E} [|X_s^{t,x} - X_{s'}^{t',x'}|^{2p}] \leq K [|t - t'|^p + |s - s'|^p + \|x - x'\|_d^{2p}].$$

11.3.2 Généralisation de la formule de Feynman-Kac

Théorème 11.5 — *Soit w une fonction de classe $C^{1,2}$ sur $[0, T] \times \mathbb{R}^d$ et à valeurs dans $[0, T] \times \mathbb{R}^k$ telle que w soit solution de*

$$(1) \quad \begin{cases} \partial_t w(t, x) + L_t w(t, x) + f(t, x, w(t, x), \nabla_x w(t, x), \sigma(t, x)) = 0, \\ w(T, x) = g(x), \end{cases}$$

où ∇_x est la Jacobienne, i.e. la matrice $\partial/\partial x_i$. On suppose que

$$|w(t, x)| + |\nabla_x w(t, x) \sigma(t, x)| \leq K(1 + |x|).$$

Alors

$$\begin{aligned} w(t, x) &= Y_t^{t,x} \\ &= \mathbb{E} \left[g(X_T^{t,x}) + \int_t^T f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x}) ds \right], \end{aligned}$$

où $(Y^{t,x}, Z^{t,x})$ est la solution de $(\star\star)$. De plus,

$$Z_u^{t,x} = \nabla_x w(u, X_u^{t,x}) \sigma(u, X_u^{t,x}).$$

Nota — L'équation

$$\frac{\partial}{\partial t} + L_t + f(t, x, u, \nabla_u) = 0$$

est appelée **équation semi-linéaire**.

11.4 Applications des équations différentielles stochastiques rétrogrades aux solutions de viscosité d'une famille d'équations aux dérivées partielles non linéaires du second ordre

11.4.1 Équation projective (*forward*)

Soient les deux fonctions continues en (t, x) et lipschitziennes en x uniformément en t suivantes :

- $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$;
- $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$.

Soit $(B_t)_t$ un mouvement brownien de dimension d . Pour tout $x \in \mathbb{R}^d$, $t \in (0, T]$, $(X_s^{t,x})$ est la solution de (1) partant de x à l'instant t :

$$X_s^{t,x} = x + \int_t^s b(u, X_u^{t,x}) du + \int_t^s \sigma(u, X_u^{t,x}) dB_u .$$

11.4.2 Équation rétrogradée (*backward*)

Soient :

- $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ continue et vérifiant

$$|g(x)| \leq K(1 + |x|^p) \quad p \geq \frac{1}{2} ;$$

- $f : [0, T] \times \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}^{k \times d} \rightarrow \mathbb{R}^k$ continue et vérifiant

$$|f(s, x, y, z) - f(s, x, y', z')| \leq K(|y - y'| + \|z - z'\|)$$

et

$$|f(t, x, y, z)| \leq K(1 + |x|^p + |y| + \|z\|) .$$

Soit $\zeta = g(X_T^{t,x})$ où le $X^{t,x}$ est celui solution de (1). Soit

$$Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x}) du - \int_s^T Z_u^{t,x} dB_u .$$

On considère l'EDP

$$\begin{cases} \frac{\partial u_i}{\partial t}(t, x) + L_t u_i(t, x) + f_i(t, x, u(t, x), (\nabla u \sigma)(t, x)) = 0 , \\ u(T, x) = g(x) , \quad x \in \mathbb{R}^d , \end{cases}$$

où les notations sont les mêmes qu'en page 111 et où u est une fonction de $[0, T] \times \mathbb{R}^d$ dans \mathbb{R}^k .

On a donc

$$\left(\frac{\partial}{\partial t} + L_t \right) = -f_i ,$$

avec f_i non linéaire.

On fait l'hypothèse « technique » que $f_i(t, x, y, z)$ ne dépend que de la i^e ligne de la matrice z :

$$f_i(t, x, u(t, x), (\nabla u \sigma)(t, x)) = f_i(t, x, u(t, x), \underbrace{(\nabla u \sigma)}_{\nabla u_i \sigma} \cdot (t, x)) .$$

Définition 11.6 — *Nous avons :*

a) $u \in \mathcal{C}([0, T] \times \mathbb{R}^d, \mathbb{R}^k)$ est une **sous-solution de viscosité** de l'équation (\star) suivante :

$$(\star) \quad \begin{cases} \left(\frac{\partial}{\partial t} + L_t \right) u = -f_i \\ u(T, x) = g(x) \end{cases}$$

si

- $u_i(T, x) \leq g_i(x), \quad x \in \mathbb{R}^d, 1 \leq i \leq d,$
- $\forall i = 1, \dots, k, \quad \phi : C^{1,2}([0, T] \times \mathbb{R}^d) \rightarrow \mathbb{R}, \quad \text{pour } (t, x) \in [0, T] \times \mathbb{R}^d$
maximum local de $u_i - \phi$,

on a

$$-\frac{\partial \phi}{\partial t}(t, x) - L_t \phi(t, x) - f_i(t, x, u(t, x), (\nabla \phi \sigma)(t, x)) \leq 0;$$

b) $u \in \mathcal{C}([0, T] \times \mathbb{R}^d, \mathbb{R}^k)$ est une **sur-solution de viscosité** de l'équation (\star) si :

- $u_i(T, x) \geq g_i(x), \quad x \in \mathbb{R}^d, 1 \leq i \leq d,$
- $\forall i = 1, \dots, k, \quad \phi : C^{1,2}([0, T] \times \mathbb{R}^d) \rightarrow \mathbb{R}, \quad \text{pour } (t, x) \in [0, T] \times \mathbb{R}^d$
minimum local de $u_i - \phi$,

on a

$$-\frac{\partial \phi}{\partial t}(t, x) - L_t \phi(t, x) - f_i(t, x, u(t, x), (\nabla \phi \sigma)(t, x)) \geq 0;$$

b) $u \in \mathcal{C}([0, T] \times \mathbb{R}^d, \mathbb{R}^k)$ est une **solution de viscosité** de (\star) si c'est une sur-solution et une sous-solution de viscosité de (\star) .

Théorème 11.6 — $u(t, x) = Y_t^{t, x}$ est une solution de viscosité de (\star) . $Y_t^{t, x}$ est déterministe et est une fonction continue de (t, x) .

12

Statistique des diffusions

12.1 Introduction

Soit

$$\begin{cases} d\zeta_t = b(t, \zeta_t) dt + \sigma(t, \zeta_t) dB_t, \\ \zeta_0 = \eta \end{cases}$$

sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Théorème 12.1 (A) — Soient les hypothèses suivantes :

- (i) $b(t, x)$ et $\sigma(t, x)$ sont continues sur $[0, +\infty[\times \mathbb{R}$;
- (ii) η est une v.a. \mathcal{F}_0 -mesurable et $\mathbb{P}(|\eta| < \infty) = 1$;
- (iii) conditions de Lipschitz locales : $\forall T > 0, \forall N > 0, \exists L_{T, N} tq \forall t \in [0, T], \forall x, \forall y,$

$$|x| \leq N \text{ et } |y| \leq N \quad \Longrightarrow \quad \begin{cases} |b(t, x) - b(t, y)| \leq L_{T, N} |x - y| \\ |\sigma(t, x) - \sigma(t, y)| \leq L_{T, N} |x - y| \end{cases}$$

- (iv) croissance sous-linéaire : $\forall T > 0, \exists K_T, \forall t \in [0, T], \forall x \in \mathbb{R},$

$$|b(t, x)| + |\sigma(t, x)| \leq K_T(1 + |x|) ;$$

- (v) $\mathbb{E}(\eta^{2m}) < \infty$ pour un $m \geq 1$.

Sous les conditions (ii), (iii) et (iv), l'EDS admet un processus solution $(\zeta_t, t \geq 0)$ défini sur Ω et tel que :

- a) $\zeta_0 = \eta$ et la trajectoire $(\zeta_t, t \geq 0)$ est p.s. continue ;
- b) $(\zeta_t)_t$ est \mathcal{F}_t -adapté (i.e. est solution forte) ;
- c) si ζ^1 et ζ^2 sont deux processus solutions vérifiant a) et b), alors

$$\mathbb{P}(\forall t \geq 0, \zeta_t^1 = \zeta_t^2) = 1 .$$

Si, de plus, (v) est vérifiée, alors $\forall t \geq 0, \mathbb{E}(\zeta_t^{2m}) < \infty$.

Théorème 12.2 (B) — Soit l'EDS unidimensionnelle réelle

$$\begin{cases} d\zeta_t = b(\zeta_t) dt + \sigma(\zeta_t) dB_t, \\ \zeta_0 = \eta. \end{cases}$$

Soient les hypothèses suivantes :

- (i) η est \mathcal{F}_0 -mesurable et $|\eta| < \infty$ p.s. ;
- (ii) b est lipschitzienne sur \mathbb{R} , σ est hölderienne d'exposant $\alpha \in [\frac{1}{2}, 1]$: $\exists K > 0, \forall x, y \in \mathbb{R}$,

$$\begin{cases} |b(x) - b(y)| \leq K \cdot |x - y| \\ |\sigma(x) - \sigma(y)| \leq |x - y|^\alpha; \end{cases}$$

- (iii) $\mathbb{E}(\eta^2) < \infty$.

Alors on a le même résultat qu'au théorème précédent (et sous (iv), $\mathbb{E}(\zeta_t^2) < \infty, \forall t$).

Théorème 12.3 — Sous les hypothèses des théorèmes A et B, la loi de probabilité \mathbb{P}^T du processus solution $(\zeta_t, t \geq 0)$ ne dépend que des fonctions $b(t, x), \sigma(t, x)$ et de la loi μ de la v.a. η .

Théorème 12.4 — Soient les hypothèses suivantes :

- (i) $\forall \theta$, les fonctions $(t, x) \rightarrow b(t, x)$ et $(t, x) \rightarrow \sigma(t, x)$ satisfont les hypothèses des théorèmes A et B;
- (ii) $\mathbb{P}(\sigma(t, \zeta_t^\theta) > 0, \forall t \in [0, T]) = 1, \forall \theta \in \Theta$.

Alors $\forall \theta, \theta' \in \Theta$, les lois \mathbb{P}_θ^T et $\mathbb{P}_{\theta'}^T$ sont équivalentes et

$$\frac{d\mathbb{P}_\theta^T}{d\mathbb{P}_{\theta'}^T}(x) = \exp \left[\int_0^T \frac{b(t, X_t, \theta) - b(t, X_t, \theta')}{\sigma^2(t, X_t)} dX_t - \frac{1}{2} \int_0^T \frac{b^2(t, X_t, \theta) - b^2(t, X_t, \theta')}{\sigma^2(t, X_t)} dt \right].$$

La **fonction de vraisemblance** associée à l'observation $(\zeta_t, t \geq 0)$ est

$$\theta \longmapsto L_T(\theta) = \exp \left[\int_0^T \frac{b(s, \zeta_s, \theta)}{\sigma^2(s, \zeta_s)} d\zeta_s - \frac{1}{2} \int_0^T \frac{b^2(s, \zeta_s, \theta)}{\sigma^2(s, \zeta_s)} ds \right].$$

Définition 12.1 — L'estimateur $\hat{\theta}$ de θ_0 est dit **faiblement consistant** si

$$\hat{\theta} \xrightarrow{\mathbb{P}} \theta_0 \quad (T \rightarrow +\infty).$$

Définition 12.2 — L'estimateur $\hat{\theta}$ de θ_0 est dit **fortement consistant** si

$$\hat{\theta} \xrightarrow{p.s.} \theta_0 \quad (T \rightarrow +\infty).$$

Théorème 12.5 — Soit

$$M_t = \int_0^t H_s dB_s$$

avec $(H_t)_t$ processus progressivement mesurable. Alors :

(i) si

$$\begin{aligned} \langle M \rangle_\infty &= \int_0^\infty H_s^2 ds \\ &= +\infty \quad p.s. \end{aligned}$$

alors

$$\frac{M_t}{\langle M \rangle_t} \xrightarrow{p.s.} 0 \quad (t \rightarrow +\infty);$$

(ii) si

$$\frac{\langle M \rangle_T}{\phi(T)} \xrightarrow{\mathbb{P}} \sigma^2 \quad (T \rightarrow +\infty),$$

où ϕ est une fonction déterministe, croissante, tendant vers $+\infty$ quand T tend vers $+\infty$, alors

$$\frac{M_T}{\sqrt{\phi(T)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad (T \rightarrow +\infty).$$

(Si $\sigma^2 = 0$, alors $\frac{M_T}{\sqrt{\phi(T)}} \xrightarrow{\mathbb{P}} 0$).

Remarque — Ce résultat reste valable pour le cas multidimensionnel.

12.2 Processus d'Ornstein-Uhlenbeck

Soit

$$\begin{cases} d\zeta_t = \theta_0 \zeta_t dt + dB_t, \\ \zeta_0 = x_0. \end{cases}$$

L'estimateur du maximum de vraisemblance (EMV) est

$$\begin{aligned} \hat{\theta}_T &= \frac{\int_0^T \zeta_s d\zeta_s}{\int_0^T \zeta_s^2 ds} \\ &= \theta_0 + \frac{\int_0^T \zeta_s dB_s}{\int_0^T \zeta_s^2 ds}. \end{aligned}$$

Proposition 12.1 — Si $\theta_0 < 0$, alors

$$\frac{1}{T} \int_0^T \zeta_s^2 ds \xrightarrow{\mathcal{L}^2} \frac{1}{2|\theta_0|} \quad (T \rightarrow +\infty).$$

Corollaire 12.1 — $\widehat{\theta}_T$ est fortement consistant et

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2|\theta_0|) \quad (T \rightarrow +\infty).$$

Proposition 12.2 — Nous avons :

1) si $\theta_0 = 0$,

$$T \cdot \widehat{\theta}_T \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_u dB_u}{\int_0^1 B_u^2 du},$$

où $(B_u)_u$ est le brownien ;

2) si $\theta_0 > 0$, on pose

$$m_T(\theta_0) = \frac{e^{2\theta_0} - 1}{2\theta_0}$$

et

$$Z = x_0 + \int_0^{+\infty} e^{-\theta_0 s} dB_s ;$$

alors

$$\frac{1}{m_T(\theta_0)} \int_0^T \zeta_s^2 ds \xrightarrow{\mathcal{L}^1} Z^2 \quad (T \rightarrow +\infty)$$

$$m_T(\theta_0)^{\frac{1}{2}} (\widehat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \frac{U}{Z}$$

et

$$\left(\int_0^T \zeta_s^2 ds \right)^{\frac{1}{2}} (\widehat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

où $(U, Z) \rightsquigarrow \mathcal{N}(0, 1) \otimes \mathcal{N}(x_0, \frac{1}{2\theta_0})$.

12.3 Markov et les diffusions

On étudie

$$\begin{cases} d\zeta_t = b(\zeta_t) dt + \sigma(\zeta_t) dB_t, \\ \zeta_0 = \eta, \end{cases}$$

avec :

- b et σ de classe \mathcal{C}^1 sur \mathbb{R} ;
- $\exists K \geq 0, \forall x \in \mathbb{R}, b^2(x) + \sigma^2(x) \leq K(1 + x^2)$.

Théorème 12.6 — $(\zeta_t, t \geq 0)$ est un processus de Markov de probabilité de transition homogène dans le temps, ne dépendant que de b et σ , i.e. :

(i) *propriété de Markov* : $\forall A \in \mathcal{B}(\mathbb{R}), 0 \leq s < t,$

$$\mathbb{P}(\zeta_t \in A \mid \mathcal{F}_s) = \mathbb{P}(\zeta_t \in A \mid \zeta_s);$$

(ii) *propriété d'homogénéité* :

$$\mathbb{P}(\zeta_t \in A \mid \zeta_s = x) = P_{t-s}(x; A),$$

où $P_t(x; dy)$ est la probabilité de transition.

Notation — Nous notons :

$$P_t(x; dy) = p_t(x, y) dy.$$

$p_t(x, y)$ est la densité de transition.

Proposition 12.3 — *Nous avons :*

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} [\zeta_{t+h} - \zeta_t \mid \zeta_t = x] &= \lim_{h \rightarrow 0} \frac{1}{h} \int (y - x) p_h(x, y) dy \\ &= b(x). \end{aligned}$$

Cette quantité est appelée **moyenne infinitésimale**.

Proposition 12.4 — *Nous avons :*

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} [(\zeta_{t+h} - \zeta_t)^2 \mid \zeta_t = x] &= \lim_{h \rightarrow 0} \frac{1}{h} \int (y - x)^2 p_h(x, y) dy \\ &= \sigma^2(x). \end{aligned}$$

Cette quantité est appelée **variance infinitésimale**.

Définition 12.3 — *Soit $f \in C_b(\mathbb{R})$.*

$$\begin{aligned} P_t f(x) &= \mathbb{E} [f(\zeta_t^x)] \\ &= \mathbb{E} [f(\zeta_t) \mid \zeta_0 = x]. \end{aligned}$$

On appelle **générateur infinitésimal** du processus $(\zeta_t)_t$ l'opérateur

$$L f(x) = \lim_{t \rightarrow 0} \frac{1}{t} [P_t f(x) - f(x)]$$

lorsque cette limite existe.

Théorème 12.7 — *Si $f \in C_K^2(\mathbb{R})$, alors*

$$L f(x) = \frac{1}{2} \sigma^2(x) f''(x) + b(x) f'(x).$$

Définition 12.4 — μ , loi de probabilité sur \mathbb{R} , est une **distribution stationnaire** pour $(\zeta_t)_t$ si

$$\zeta_0 \rightsquigarrow \mu \implies \forall t, \zeta_t \rightsquigarrow \mu.$$

Théorème 12.8 — Nous avons :

$$\mu \text{ distribution stationnaire} \implies \forall f \in C_K^2(\mathbb{R}), \int Lf(x) d\mu(x) = 0.$$

Théorème 12.9 — On suppose $\sigma^2(x) > 0, \forall x \in \mathbb{R}$, σ de classe C^2 , b de classe C^1 . Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ de classe C^2 , positive et telle que $\int_{\mathbb{R}} h(x) dx = 1$. Alors

$$h(x) dx = \mu(x) dx$$

est une distribution stationnaire pour $(\zeta_t)_t$ ssi

$$\left(\frac{1}{2}h\sigma^2\right)'' - (hb)' = 0.$$

Définition 12.5 — Le **facteur intégrant** du processus est

$$s(x) = \exp\left[-2 \int^x \frac{b(u)}{\sigma^2(u)} du\right].$$

Définition 12.6 — Le **facteur d'échelle** du processus est

$$S(x) = \int^x s(u) du.$$

Théorème 12.10 — Si

$$\begin{aligned} \lim_{x \rightarrow +\infty} S(x) &= +\infty, \\ \lim_{x \rightarrow -\infty} S(x) &= -\infty \end{aligned}$$

et si

$$M = \int \frac{dx}{\sigma^2(x)s(x)} < \infty,$$

alors

$$\mu(x) = \frac{dx}{M\sigma^2(x)s(x)}$$

est une distribution stationnaire.

Soit $I = (l, r)$, $-\infty \leq l < r \leq +\infty$, tel que :

- $\sigma^2(x) > 0, \forall x \in I$;
- b de classe C^1 sur I ;
- σ de classe C^2 sur I .

Définition 12.7 — Soient $x, y \in I$. Le **temps d'atteinte** de y est défini par

$$T_{x, y} = \inf\{ t \geq 0, \zeta_t^x = y \}.$$

Le **temps d'explosion** est défini par

$$e_x = \inf\{ t \geq 0, \zeta \notin (l, r) \}.$$

Proposition 12.5 —

$$\forall x, y \in I, \quad \mathbb{P}(T_{x, y} < \infty) > 0.$$

Théorème 12.11 — Soient $l < a < x < b < r$ et

$$\begin{aligned} T &= \inf\{ t \geq 0, \zeta_t^x = a \text{ ou } b \} \\ &= T_{x, a} \wedge T_{x, b}. \end{aligned}$$

Alors

$$\begin{aligned} \mathbb{P}(T < \infty) &= 1 \\ \mathbb{P}(\zeta_T^x = a) &= \mathbb{P}(T_{x, a} < T_{x, b}) \\ &= \frac{S(b) - S(x)}{S(b) - S(a)} \\ \mathbb{P}(\zeta_T^x = b) &= \mathbb{P}(T_{x, b} < T_{x, a}) \\ &= \frac{S(x) - S(a)}{S(b) - S(a)}. \end{aligned}$$

Théorème 12.12 — Nous avons :

$$\begin{cases} S(l^+) = -\infty \\ S(r^-) = +\infty \end{cases} \implies \forall x, y \in I, \mathbb{P}(T_{x, y} < \infty) = 1.$$

Le processus est alors dit **récurrent** sur I . De plus,

$$\mathbb{P}(e_x = +\infty) = 1.$$

Remarque — Nous avons :

$$\begin{aligned} S(l^+) = -\infty &\iff \int_l^r s(u) \, du = -\infty, \\ S(r^-) = +\infty &\iff \int_l^r s(u) \, du = +\infty. \end{aligned}$$

Théorème 12.13 — Soient $l < a < x < b < r$ et $T = T_{x,a} \wedge T_{x,b}$. Soit u la fonction de classe C^2 et définie sur I par

$$\begin{cases} Lu = -1, \\ u(a) = u(b) = 0. \end{cases}$$

Alors

$$\begin{aligned} u(x) &= \mathbb{E}(T) \\ &= 2 \left\{ \frac{S(x) - S(a)}{S(b) - S(a)} \int_x^b \frac{S(b) - S(u)}{\sigma^2(u)s(u)} \, du + \frac{S(b) - S(x)}{S(b) - S(a)} \int_a^x \frac{S(u) - S(a)}{\sigma^2(u)s(u)} \, du \right\}. \end{aligned}$$

Théorème 12.14 — Si

$$\begin{aligned} S(l^+) &= -\infty, \\ S(r^-) &= +\infty \end{aligned}$$

et si

$$M = \int_l^r \frac{du}{\sigma^2(u)s(u)} < \infty,$$

alors $\forall x, y \in I$

$$\mathbb{E}(T_{x,y}) < \infty.$$

Le processus est alors dit **récurrent positif** sur I .

$$m(u) \, du = \frac{du}{\sigma^2(u)s(u)}$$

est appelée **mesure de vitesse**.

Théorème 12.15 — Si un processus est récurrent positif, alors il admet une unique distribution stationnaire, qui est donnée par

$$\pi(x) \, dx = \frac{m(x)}{M} \mathbb{1}_{\{x \in I\}} \, dx.$$

Théorème 12.16 — Nous avons :

1) quelle que soit la loi initiale pour $(\zeta_t, t \geq 0)$,

$$\zeta_t \xrightarrow{\mathcal{L}} \pi \quad (t \rightarrow +\infty);$$

2) quelle que soit la loi initiale pour $(\zeta_t, t \geq 0)$,

$$\frac{1}{T} \int_0^T f(\zeta_s) \, ds \xrightarrow{p.s.} \int_I f(x) \pi(x) \, dx \quad (T \rightarrow +\infty)$$

dès que $\int_I |f(x)| \pi(x) \, dx < \infty$.

12.3.1 Étude des estimateurs du maximum de vraisemblance

On étudie les solutions de l'équation

$$l_T(\widehat{\theta}_T) = \sup_{\theta \in \Theta} l_T(\theta).$$

Tout d'abord, on cherche les hypothèses pour que

$$\mathbb{P}(|\widehat{\theta}_T - \theta_0| > h) \rightarrow 0 \quad (T \rightarrow +\infty).$$

On note $K(\cdot)$ les hypothèses suivantes :

(K1) Θ est un compact de \mathbb{R}^p ;

(K2) $\theta \mapsto l_T(\theta)$ admet une version continue sur Θ ;

(K3) il existe une v.a. Z_T et une fonction $\beta(\eta)$ telle que $\forall \theta, \theta' \in \Theta$,

$$|\theta - \theta'| \leq \eta \Rightarrow \frac{1}{T} |l_T(\theta) - l_T(\theta')| \leq \beta(\eta) Z_T,$$

avec $\beta(\eta) \rightarrow 0$ ($\eta \rightarrow \theta$) et Z_T convergeant en proba quand T tend vers l'infini ;

(K4) soit

$$K(\theta_0, \theta) = \int_l^r \left(\frac{b(u, \theta_0) - b(u, \theta)}{\sigma(u)} \right)^2 \pi_{\theta_0}(u) \, du ;$$

alors

$$\theta \neq \theta_0 \Leftrightarrow K(\theta_0, \theta) < \infty \quad (\text{hypothèse d'identifiabilité}) ;$$

(K5) $\theta \mapsto K(\theta_0, \theta)$ est continue.

Proposition 12.6 — On a, sous (K4),

$$\frac{1}{T} [l_T(\theta_0) - l_T(\theta)] \xrightarrow{p.s.} \frac{1}{2} K(\theta_0, \theta) \quad (T \rightarrow \infty).$$

Nota — La fonction $\theta \mapsto -\frac{1}{T} l_T(\theta)$ est une **fonction de contraste**.

Théorème 12.17 — Sous les hypothèses $K(i)$, $i = 1, \dots, 5$, on a $\forall h > 0$,

$$\mathbb{P}(|\hat{\theta}_T - \theta_0| > h) \longrightarrow 0 \quad (T \rightarrow \infty).$$

Problématique — Il s'agit de vérifier que $\theta \mapsto l_T(\theta)$ admet une version continue en θ .

$$\theta \mapsto \int_0^T \frac{b(\theta, \zeta_s)}{\sigma^2(\zeta_s)} d\zeta_s - \frac{1}{2} \int_0^T \frac{b^2(\theta, \zeta_s)}{\sigma^2(\zeta_s)} ds.$$

Le second terme peut être traité par le théorème classique de Lebesgue. Le premier vaut

$$\int_0^T \frac{b(\theta, \zeta_s)}{\sigma^2(\zeta_s)} \left[b(\theta_0, \zeta_s) ds + \sigma(\zeta_s) dW_s \right].$$

Question :

$$\theta \mapsto M_T(\theta) = \int_0^T \phi(\theta, t, \zeta_t) dW_t$$

admet-elle une version continue ?

Théorème 12.18 (Kolmogorov) — $\exists \gamma > 0$, $\epsilon > 0$, $c > 0$ tels que $\forall \theta, \theta'$,

$$\mathbb{E} \left(|M_T(\theta) - M_T(\theta')|^\gamma \right) \leq c \cdot |\theta - \theta'|^{p+\epsilon},$$

où $\Theta \subset \mathbb{R}^p$. Par conséquent, $\theta \mapsto M_T(\theta)$ admet une version continue en θ .

On fait les hypothèses supplémentaires suivantes :

(K6) θ_0 , vraie valeur du paramètre, appartient à $\overset{\circ}{\Theta}$;

(K7) les fonctions $b'_{\theta_i}(x, \theta)$ et $b''_{\theta_i \theta_j}(x, \theta)$ sont définies et continues sur $(l, r) \times \overset{\circ}{\Theta}$ et

$$\frac{\partial l_T}{\partial \theta_i}(\theta_0) = \int_0^T \frac{b'_{\theta_i}(\zeta_s, \theta_0)}{\sigma^2(\zeta_s)} [d\zeta_s - b(\zeta_s, \theta_0) ds],$$

$$\frac{\partial^2 l_T}{\partial \theta_i \partial \theta_j}(\theta_0) = \int_0^T \frac{b'_{\theta_i \theta_j}(\zeta_s, \theta_0)}{\sigma^2(\zeta_s)} [d\zeta_s - b(\zeta_s, \theta_0) ds] - \int_0^T \frac{b'_{\theta_i}(\zeta_s, \theta_0) b'_{\theta_j}(\zeta_s, \theta_0)}{\sigma^2(\zeta_s)} ds;$$

(K8) l'**information de Fischer** est $\mathcal{I}(\theta_0) = (\mathcal{I}_{ij}(\theta_0))_{ij}$ avec

$$\mathcal{I}_{ij}(\theta_0) = \int_l^r \frac{b'_{\theta_i}(u, \theta_0) b'_{\theta_j}(u, \theta_0)}{\sigma^2(u)} \pi_{\theta_0}(u) du;$$

$\mathcal{I}(\theta_0)$ est bien définie et inversible;

(K9) soit

$$\mathcal{J}_{ij}(\theta_0) = \int_l^r \frac{b''_{\theta_i \theta_j}(u, \theta_0)}{\sigma^2(u)} \pi_{\theta_0}(u) du;$$

(K10) on a

$$\sup_{|\alpha|} \frac{1}{T} |l''_{T,\theta_i, \theta_j}(\theta_0) - l''_{T,\theta_i, \theta_j}(\theta_0 + \alpha)| \xrightarrow{\mathbb{P}} 0 \quad (T \rightarrow \infty).$$

Remarque — Si $\Theta \subset \mathbb{R}$, (K8) équivaut à

$$\mathcal{I}(\theta_0) = \int_l^r \frac{b'_\theta{}^2(u, \theta_0)}{\sigma^2(u)} \pi_{\theta_0}(u) du.$$

Proposition 12.7 — On a :

1)

$$\frac{1}{\sqrt{T}} \left(\frac{\partial}{\partial \theta_i} l_T(\theta_0) \right)_{i=1, \dots, p} \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \mathcal{I}(\theta_0)) \quad (T \rightarrow \infty);$$

2)

$$\frac{1}{\sqrt{T}} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_T(\theta_0) \right)_{1 < i, j < p} \xrightarrow{\mathbb{P}} -\mathcal{I}(\theta_0) \quad (T \rightarrow \infty).$$

Théorème 12.19 — Sous K(i), $i=1, \dots, 10$,

$$\sqrt{T} (\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \mathcal{I}(\theta_0)) \quad (T \rightarrow \infty).$$

12.4 Estimateurs empiriques

On suppose $(\zeta_t, t \geq 0)$ récurrent positif.

Théorème 12.20 (Ergodicité) — Si $f : (l, r) \rightarrow \mathbb{R}$, borélienne et telle que $\int_I |f| d\pi < \infty$, alors

$$\frac{1}{T} \int_0^T f(\zeta_s) ds \xrightarrow{p.s.} \int_I f(x) \pi(x, \theta_0) dx,$$

quelle que soit la loi de ζ_0 .

Théorème 12.21 (Convergence en loi) — Si $f : (l, r) \rightarrow \mathbb{R}$, borélienne et telle que d'une part $\int_I |f| d\pi < \infty$, d'autre part $\int_I f d\pi = 0$, alors

$$\frac{1}{\sqrt{T}} \int_0^T f(\zeta_s) ds \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_{\theta_0}(f)),$$

à condition que $V_{\theta_0}(f)$ soit finie.

Soit

$$Af(x, \theta_0) = \int_l^x f(u) \pi(u, \theta_0) du .$$

Alors

$$V_{\theta_0}(f) = 4M(\theta_0) \int_I s(x, \theta_0) A^2 f(x, \theta_0) dx$$

avec

$$M(\theta_0) = \int_I m(x, \theta_0) dx ,$$

$$m(x, \theta_0) = \frac{1}{\sigma^2(x)s(x, \theta_0)}$$

et

$$s(x, \theta_0) = \exp \left[-2 \int_{x_0}^x \frac{b(u, \theta_0)}{\sigma^2(u)} du \right] .$$

Corollaire 12.2 — Soient $f_1, \dots, f_k : I \rightarrow \mathbb{R}$, continues et telles que d'une part $\int_I |f_i| \pi < \infty$, d'autre part $\int_I f_i d\pi = 0$. Alors

$$\frac{1}{\sqrt{T}} \left(\int_0^T f_i(\zeta_s) ds \right)_{i=1, \dots, k} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, (V_{\theta_0}(f_i, f_j))_{1 \leq i, j \leq k} \right) ,$$

à condition que $V_{\theta_0}(f_i)$ soit finie, pour $i = 1, \dots, k$.

Théorème 12.22 — Soit $f : I \rightarrow \mathbb{R}$, continue et telle que $|f(x)| \leq K(1 + |x|^\gamma)$ avec

$$\int_I |x|^\gamma \pi_{\theta_0}(x) dx < \infty .$$

Si

$$\int_l s(x, \theta_0) dx \left(\int_l^x (1 + |u|^\gamma) \pi_{\theta_0}(u) du \right)^2 < \infty$$

et si

$$\int_x^r s(x, \theta_0) dx \left(\int_x^r (1 + |u|^\gamma) \pi_{\theta_0}(u) du \right)^2 < \infty ,$$

alors

$$V_{\theta_0}(f) < \infty .$$

Soit $\hat{\theta}_T$ l'EMV et $\tilde{\theta}$ un autre estimateur de θ . Posons

$$\hat{\tilde{\theta}}_T = \tilde{\theta}_T - \frac{l'(\tilde{\theta})}{l''(\tilde{\theta})}$$

(méthode de Newton au premier pas). Si $\hat{\theta}_T$ est consistant et si :

1.
$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0)) ;$$

2. $\hat{\theta}_T \longrightarrow \theta_0 ;$

3. $\sqrt{T}(\tilde{\theta}_T - \theta_0)$ converge en loi,

alors

$$\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) \xrightarrow{\mathbb{P}} 0 \quad (T \rightarrow \infty)$$

et donc

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0)) .$$

Cinquième partie

**MODÈLE LINÉAIRE
GÉNÉRALISÉ**

Introduction

13.1 Modèle linéaire classique

Le vecteur Y des observations a n composantes qui sont indépendamment distribuées, et de moyenne μ . La part systématique du modèle est la spécification de μ en fonction de paramètres β_1, \dots, β_p :

$$\mu = \sum_{j=1}^p x_j \beta_j .$$

i.e.

$$\begin{aligned} \mathbb{E}(y_i) &= \mu_i \\ &= \sum_{j=1}^p x_{ij} \beta_j , \end{aligned}$$

où x_{ij} est la valeur de la j^e covariable pour l'observation i . L'erreur du modèle suit une $\mathcal{N}(0, \sigma^2)$.

Le vocabulaire est le suivant :

- **la composante aléatoire** : les composantes de Y ont des distributions normales indépendantes d'espérances μ_i et de variance commune σ^2 ;
- **la composante systématique** : les covariables x_1, \dots, x_p engendrent un prédicteur linéaire η donné par :

$$\eta = \sum_{j=1}^p x_j \beta_j$$

- **le lien** entre composantes aléatoires et systématique est

$$\eta = \mu .$$

13.2 Modèle linéaire général

Soit y_{ij} , $j = 1, \dots, n$ les n observations faites sur le i^e sujet ; est associé au vecteur des observations un vecteur de p covariables x_{ijk} , $k = 1, \dots, p$. On suppose que les y_{ij} sont les réalisations de v.a. Y_{ij} , suivant le modèle

$$Y_{ij} = \beta_1 x_{ij1} + \dots + \beta_p x_{ijp} + \epsilon_{ij} .$$

Les erreurs sont ici corrélées. Si l'on note $\sigma^2 V$ la matrice bloc-diagonale composée de $n \times n$ blocs $\sigma^2 V_0$, chacun représentant la matrice de variance-covariance du vecteur de mesures chez un sujet, le modèle s'écrit

$$Y \rightsquigarrow \mathcal{N}(X\beta, \sigma^2 V) . \quad (13.1)$$

13.2.1 Estimation par les moindres carrés ordinaires

L'estimateur des moindres carrés ordinaire $\hat{\beta}$ minimise la forme quadratique

$$(y - X\beta)^t (y - X\beta) .$$

Il est égal à

$$\hat{\beta} = (X^t X)^{-1} X^t y$$

et

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^t X)^{-1} X^t V X (X^t X)^{-1} .$$

13.2.2 Estimation par les moindres carrés pondérés

L'estimateur des moindres carrés pondérés (*weighted least-squares estimator*) de β , qui utilise une matrice symétrique de pondération W , est la valeur $\tilde{\beta}_W$ qui minimise la forme quadratique

$$(y - X\beta)^t W (y - X\beta) .$$

Le résultat explicite est

$$\tilde{\beta}_W = (X^t W X)^{-1} X^t W y . \quad (13.2)$$

Cet estimateur est **sans biais**, quel que soit le choix de W . Sa variance vaut

$$\text{Var}(\tilde{\beta}_W) = \sigma^2 \{ (X^t W X)^{-1} X^t W \} V \{ W X (X^t W X)^{-1} \} . \quad (13.3)$$

Si $W = I$, matrice d'identité, alors on retrouve l'estimation par les moindres carrés ordinaires. Si $W = V^{-1}$, l'estimateur devient

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} y \quad (13.4)$$

et

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^t V^{-1} X)^{-1} .$$

La notation « chapeau » anticipe sur le fait que $\hat{\beta}$ est l'estimateur du maximum de vraisemblance de β sous l'hypothèse de normalité du modèle (13.1). Cette remarque laisse suggérer que l'estimateur des moindres carrés pondérés le plus efficace est celui pour lequel $W = V^{-1}$.

13.2.3 Estimation par le maximum de vraisemblance sous l'hypothèse de normalité

On estime simultanément les paramètres d'intérêt, soient β, σ^2 et V_0 . Sous l'hypothèse de normalité (cf. (13.1)), la log-vraisemblance vaut

$$\mathcal{L}(\beta, \sigma^2, V_0) = -\frac{1}{2} \left\{ nm \log(\sigma^2) + m \log(|V_0|) + \frac{1}{\sigma^2} (y - X\beta)^t V^{-1} (y - X\beta) \right\}. \quad (13.5)$$

Pour une matrice V_0 donnée, l'estimateur du maximum de vraisemblance de β est l'estimateur des moindres carrés pondérés vu en (13.4), soit

$$\hat{\beta}(V_0) = (X^t V^{-1} X)^{-1} X^t V^{-1} y. \quad (13.6)$$

Son expression insérée dans (13.5), on obtient

$$\mathcal{L}(\hat{\beta}(V_0), \sigma^2, V_0) = -\frac{1}{2} \left\{ nm \log(\sigma^2) + m \log(|V_0|) + \frac{1}{\sigma^2} \text{RSS}(V_0) \right\},$$

où

$$\text{RSS}(V_0) = (y - X\hat{\beta})^t V^{-1} (y - X\hat{\beta}).$$

La dérivation de (13.6) par rapport à σ^2 donne l'estimateur du maximum de vraisemblance de σ^2 , toujours à V_0 fixé :

$$\hat{\sigma}^2(V_0) = \frac{\text{RSS}(V_0)}{nm}. \quad (13.7)$$

L'introduction de (13.6) et (13.7) dans (13.5) donne une log-vraisemblance réduite pour V_0 qui, à un terme constant près, vaut

$$\begin{aligned} \mathcal{L}_r(V_0) &= \mathcal{L}(\hat{\beta}(V_0), \hat{\sigma}^2(V_0), V_0) \\ &= -\frac{1}{2} \left\{ n \log(\text{RSS}(V_0)) + \log(|V_0|) \right\}. \end{aligned} \quad (13.8)$$

Finalement, la maximisation de $\mathcal{L}_r(V_0)$ donne \hat{V}_0 et par suite, au travers de (13.6) et (13.7), on obtient également $\hat{\beta} \equiv \hat{\beta}(\hat{V}_0)$ et $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\hat{V}_0)$.

En utilisant la vraisemblance pour les estimations simultanées de β, σ^2 et V_0 , la forme de la matrice X intervient explicitement dans l'estimation de σ^2 et V_0 . Une conséquence de ceci est que, si nous supposons une forme incorrecte pour X , nous n'obtiendrons pas d'estimateurs consistants pour σ^2 et V_0 . Aussi, une stratégie est d'élaborer un modèle complet pour les profils des réponses moyennes qui incorpore la structure de covariance des données. Quand, par exemple, les données proviennent d'une expérimentation planifiée (*designed experiment*), et qu'il n'y a pas de covariable continue, il est recommandé d'introduire un paramètre séparé pour la réponse moyenne à chaque temps de contrôle du traitement, ce qui s'appelle un **modèle saturé** pour les profils de réponse moyenne. Ceci garantit des estimateurs consistants de la structure de covariance.

Cette stratégie n'est pas toujours praticable. En particulier, lorsqu'il existe une ou plusieurs covariables continues, nous devons décider d'introduire cette (ces) covariable(s) sous forme d'un effet linéaire, ou quadratique, ou sous une autre forme non-linéaire. Dans ce cas, le concept de modèle saturé ne tient plus.

Même s'il est praticable, le modèle saturé pose un autre problème. Pour g traitements et n temps d'observations, il requiert $p = n \times g$ paramètres, et si ce nombre est relativement important, les estimateurs du maximum de vraisemblance pour σ^2 et V_0 seront sérieusement biaisés. Par exemple, nous savons que lorsque $V_0 = I$, un estimateur sans biais de σ^2 exige

un diviseur égal à $(nm - p)$, plutôt que le diviseur nm vu en (13.7) — ce problème étant encore davantage exacerbé par la structure d'autocorrélation des données.

Aussi est-il nécessaire d'utiliser une matrice X présentant un grand nombre de colonnes pour obtenir des estimateurs consistants de la structure de covariance, alors même qu'une estimation non biaisée exige un faible nombre de colonnes pour X .

Pour remédier à ce problème, nous devons considérer d'autres méthodes d'estimation. Parmi elles, la méthode du maximum de vraisemblance restreint.

13.2.3.1 Estimation par le maximum de vraisemblance restreint

La méthode du maximum de vraisemblance produit des estimateurs biaisés; par exemple, dans le modèle le plus classique, soit

$$Y \rightsquigarrow \mathcal{N}(X\beta, \sigma^2 I), \quad (13.9)$$

l'estimateur du maximum de vraisemblance de σ^2 est

$$\hat{\sigma}^2 = \frac{\text{RSS}}{nm},$$

où RSS est la somme des carrés résiduelle. Cet estimateur est biaisé; l'estimateur usuel sans biais est

$$\tilde{\sigma}^2 = \frac{\text{RSS}}{nm - p},$$

où p est le nombre d'éléments de β .

Dans cet exemple, $\tilde{\sigma}^2$ est l'**estimateur du maximum de vraisemblance restreint (REML)** (*restricted maximum likelihood estimation*) de σ^2 concernant le modèle (13.9).

Dans le cadre plus général d'un modèle

$$Y \rightsquigarrow \mathcal{N}(X\beta, \sigma^2 V), \quad (13.10)$$

l'estimateur REML est défini comme étant l'estimateur du maximum de vraisemblance basé sur une transformation linéaire du jeu de donnée : soit

$$Y^* = AY,$$

de telle sorte que la distribution de Y^* ne dépende pas de β . Un moyen est de choisir pour A la matrice qui transforme Y en résidus des moindres carrés ordinaires :

$$A = I - X(X^t X)^{-1} X^t. \quad (13.11)$$

Alors Y^* a une distribution normale multivariée, centrée et singulière, quelle que soit la valeur de β . Pour obtenir une distribution normale centrée régulière, on peut utiliser uniquement $nm - p$ lignes de la matrice A définie en (13.11).

Les estimateurs résultant pour σ^2 et V_0 ne dépendent cependant pas du choix des lignes retenues, ni non plus du choix particulier de la matrice A : toute matrice telle que $\mathbb{E}(Y^*) = 0$ pour tout β donnera la même solution.

Pour les calculs, on réabsorbe σ^2 dans V , si bien que le modèle se réécrit

$$Y \rightsquigarrow \mathcal{N}(X\beta, H), \quad (13.12)$$

où $H \equiv H(\alpha)$, avec α vecteur de paramètres. Soit A telle qu'en (13.11) et B la matrice $nm \times (nm - p)$ telle que

$$B^t B = I,$$

où I est la matrice identité de dimension $(nm - p) \times (nm - p)$. Finalement, soit

$$Z = B^t Y .$$

À α fixé, l'estimateur du maximum de vraisemblance de β est l'estimateur des moindres carrés généralisés

$$\begin{aligned} \hat{\beta} &= (X^t H^{-1} X)^{-1} X^t H^{-1} Y \\ &= GY . \end{aligned}$$

Les densités de probabilité de Y et $\hat{\beta}$ sont respectivement

$$f(y) = \frac{1}{(\sqrt{2\pi})^{nm}} \frac{1}{\sqrt{|H|}} \exp \left\{ -\frac{1}{2} (y - X\beta)^t H^{-1} (y - X\beta) \right\}$$

et

$$g(\hat{\beta}) = \frac{1}{(\sqrt{2\pi})^p} \sqrt{|X^t H^{-1} X|} \exp \left\{ -\frac{1}{2} (\hat{\beta} - \beta)^t (X^t H^{-1} X) (\hat{\beta} - \beta) \right\} .$$

On a $\mathbb{E}(Z) = 0$; de plus, Z et $\hat{\beta}$ sont indépendants, quelle que soit la valeur de β . On démontre que l'estimateur REML $\tilde{\alpha}$ maximise la log-vraisemblance

$$\mathcal{L}^*(\alpha) = -\frac{1}{2} \log |H| - \frac{1}{2} \log |X^t H^{-1} X| - \frac{1}{2} (y - X\hat{\beta})^t H^{-1} (y - X\hat{\beta}) ,$$

tandis que l'estimateur du maximum de vraisemblance $\hat{\alpha}$ maximise la log-vraisemblance

$$\mathcal{L}(\alpha) = -\frac{1}{2} \log |H| - \frac{1}{2} (y - X\hat{\beta})^t H^{-1} (y - X\hat{\beta}) .$$

Ainsi, l'algorithme du REML incorpore uniquement une modification de celui du maximum de vraisemblance.

Revenons un instant au modèle de la section précédente. Si l'on considère m unités et n observations par unité, et si $\sigma^2 V$ est une matrice bloc-diagonale faite de $n \times n$ blocs non nuls $\sigma^2 V_0$ (représentant chacun la matrice de variance-covariance des mesures faites sur une unité), alors à V_0 donnée,

$$\begin{aligned} \hat{\beta}(V_0) &= (X^t V^{-1} X)^{-1} X^t V^{-1} y \\ \text{RSS}(V_0) &= (y - X\hat{\beta}(V_0))^t V^{-1} (y - X\hat{\beta}(V_0)) \end{aligned} \tag{13.13}$$

et l'estimateur REML de σ^2 est

$$\tilde{\sigma}(V_0) = \frac{\text{RSS}(V_0)}{nm - p} , \tag{13.14}$$

où p est le nombre d'éléments de β .

L'estimateur REML de V_0 maximise la vraisemblance

$$\mathcal{L}^*(V_0) = -\frac{1}{2} m \left\{ n \log \text{RSS}(V_0) + \log (|V_0|) \right\} - \frac{1}{2} \log (|X^t V^{-1} X|) . \tag{13.15}$$

Finalement, en insérant dans (13.13) et (13.14) le résultat \tilde{V}_0 obtenu par (13.15), on obtient les estimateurs REML

$$\tilde{\beta} = \hat{\beta}(\tilde{V}_0)$$

et

$$\tilde{\sigma}^2 = \hat{\sigma}^2(\tilde{V}_0).$$

Nota — La différence entre $\mathcal{L}(V_0)$ et $\mathcal{L}^*(V_0)$ réside dans l'addition du terme $\frac{1}{2} \log(|X^t V^{-1} X|)$. La matrice $X^t V^{-1} X$ est une matrice $p \times p$. Aussi la différence entre maximum de vraisemblance ordinaire et REML est-elle importante quand p est grand.

13.2.4 Estimation robuste des écarts-types

L'idée essentielle de l'approche robuste de l'inférence concernant β est d'utiliser l'estimateur des moindres carrés généralisé $\tilde{\beta}$ défini en (13.2) par

$$\tilde{\beta} = (X^t W X)^{-1} X^t W y, \quad (13.16)$$

en conjonction avec une matrice de variance-covariance estimée

$$\hat{R}_W = \{(X^t W X)^{-1} X^t W\} \hat{V} \{W X (X^t W X)^{-1}\}, \quad (13.17)$$

où \hat{V} est consistante pour V , quelle que soit la vraie structure de covariance. Notons que dans (13.17), σ^2 a été réabsorbé dans V .

Pour l'inférence, nous procédons comme si

$$\tilde{\beta} \rightsquigarrow \mathcal{N}(\beta, \hat{R}_W). \quad (13.18)$$

Dans cette approche, on appelle W^{-1} la **matrice de covariance de travail**, afin de la distinguer de la vraie matrice de covariance V . Typiquement, nous pouvons utiliser une forme simple pour W^{-1} qui « capture » la structure qualitative de V .

Quoi qu'il en soit, un choix quelconque pour W affectera seulement l'efficacité de nos inférences concernant β , mais pas leur validité. En particulier, les intervalles de confiance et les tests d'hypothèses issus de (13.18) seront asymptotiquement corrects, quelle que soit la vraie forme de V .

Notons que les équations (13.2) et (13.3) ne changent pas si les éléments de W sont multipliés par une constante, si bien qu'il serait strictement correct de dire que W^{-1} est proportionnel à la matrice de covariance de travail.

Quand le modèle saturé n'est pas envisageable (présence d'une covariable continue), il n'est pas possible d'obtenir une expression explicite de l'estimateur REML de V_0 . Dans ce cas, on ne fait aucune hypothèse au sujet de la forme de V_0 ; on utilise une matrice X correspondant au modèle le plus élaboré que nous avons pu préparer concernant la réponse moyenne; enfin l'on obtient l'estimateur REML \hat{V}_0 *via* une maximisation numérique qui est en (13.15).

Pour des inférences robustes concernant β , on substitue dans (13.17) \hat{V} et on utilise (13.18). Si l'on désire tester des hypothèses linéaires concernant β , on peut utiliser l'approche standard du modèle linéaire général. Ainsi, si l'on désire tester l'hypothèse $Q\beta = 0$, où Q est une matrice $q \times p$ avec $q < p$, on déduit de (13.18) que

$$Q\hat{\beta}_W \rightsquigarrow \mathcal{N}(Q\beta, Q\hat{R}_W Q^t).$$

Une statistique est alors

$$T = \hat{\beta}_W^t Q^t (Q\hat{R}_W Q^t)^{-1} Q\hat{\beta}_W, \quad (13.19)$$

qui suit un $\chi^2(q)$.

Modèle linéaire généralisé

14.1 Présentation

La généralisation consiste en deux points :

- la distribution de la composante aléatoire n'est plus nécessairement normale — elle est cependant issue de la famille exponentielle — ;
- le lien devient une **fonction de lien**, *i.e.*

$$\eta = g(\mu) ,$$

avec g monotone et différentiable.

L'expression générale est

$$f_Y(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}$$

pour des fonctions spécifiques a , b et c .

Notons $l(\theta_i, \phi; y_i) = \log f(y_i; \theta_i, \phi) = l_i$ la contribution de la i^e observation à la log-vraisemblance. On a :

$$\begin{aligned} l(\theta_i, \phi; y_i) &= \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \\ \Rightarrow \frac{\partial l_i}{\partial \theta_i} &= \frac{y_i - b'(\theta_i)}{a(\phi)} \end{aligned} \tag{14.1}$$

$$\text{et } \frac{\partial^2 l_i}{\partial \theta_i^2} = -\frac{b''(\theta_i)}{a(\phi)} . \tag{14.2}$$

Théorème 14.1 —

$$\mathbb{E}\left(\frac{\partial l}{\partial \theta_0}\right) = 0 , \tag{14.3}$$

$$\mathbb{E}\left[\left(\frac{\partial l}{\partial \theta_0}\right)^2\right] = -\mathbb{E}\left(\frac{\partial^2 l}{\partial \theta_0^2}\right) . \tag{14.4}$$

où θ_0 est la vraie valeur du paramètre.

Démonstration

-1-

$$\begin{aligned} \int f(y; \theta) dy &= 1 \\ \Rightarrow 0 &= \int \frac{\partial f(y; \theta)}{\partial \theta} dy \\ &= \int \frac{\partial \log f(y; \theta)}{\partial \theta} f(y; \theta) dy \\ &= \mathbb{E}\left(\frac{\partial l}{\partial \theta}\right). \end{aligned}$$

-2-

D'après (14.1),

$$\text{Var}\left(\frac{\partial l}{\partial \theta}\right) = \mathbb{E}\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right].$$

D'autre part, en dérivant l'équation du 1,

$$\begin{aligned} 0 &= \int \frac{\partial^2 \log f(y; \theta)}{\partial \theta^2} f(y; \theta) dy + \int \frac{\partial \log f(y; \theta)}{\partial \theta} \frac{\partial f(y; \theta)}{\partial \theta} dy \\ \Rightarrow \mathbb{E}\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right] &= -\mathbb{E}\left(\frac{\partial^2 l}{\partial \theta^2}\right) \\ &= A(\theta) \quad (\text{notation}) . \end{aligned}$$

$A(\theta)$ est la **matrice d'information de Fisher**. ■

D'après (14.1) et (14.3),

$$\begin{aligned} \mathbb{E}(y_i) &= b'(\theta_i) \\ &= \mu_i \quad (\text{notation}) . \end{aligned} \tag{14.5}$$

D'après (14.2) et (14.4),

$$\text{Var}(y_i) = b''(\theta_i) a(\phi) . \tag{14.6}$$

La variance se décompose en une partie ne dépendant que de θ (et donc de la moyenne), que l'on nommera **fonction de variance** et que l'on notera $V(\mu)$, et une partie dépendant uniquement de ϕ . La fonction $a(\phi)$ est souvent de la forme

$$a(\phi) = \frac{\phi}{w} ,$$

où ϕ , noté encore σ^2 et appelé **paramètre de dispersion**, est constant sur les observations, et w un poids *a priori*, connu donc, et qui varie d'une observation à l'autre.

θ_i est une fonction de μ :

$$\theta_i = b'^{-1}[g^{-1}(\eta_i)] = h(\eta_i) = h'(\mu_i) .$$

Une fonction de lien pour laquelle $\theta = \eta$ est appelée **fonction de lien canonique**.

TABLE 14.1 — Quelques lois.

Nom	Normale	Poisson	Binomiale	Gamma	Inverse Gaussienne
Notation	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{P}(\mu)$	$\frac{1}{m}\mathcal{B}(m, \pi)$	$G(\mu, \nu)$	$IG(\mu, \sigma^2)$
ϕ	σ^2	1	$\frac{1}{m}$	ν^{-1}	σ^2
$b(\theta)$	$\frac{\theta^2}{2}$	e^θ	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-\sqrt{-2\theta}$
$\mu(\theta) = \mathbb{E}(Y; \theta)$	θ	e^θ	$\frac{e^\theta}{1 + e^\theta}$	$-\frac{1}{\theta}$	$\frac{1}{\sqrt{-2\theta}}$
$\theta(\mu)$	μ	$\log(\mu)$	$\text{logit}(\mu)$	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$
$V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

14.1.1 Les équations de vraisemblance

Si l'échantillon est composé de n observations indépendantes, alors la log-vraisemblance de l'échantillon est égale à

$$\begin{aligned} L(\beta) &= \sum_{i=1}^n \log f(y_i; \theta_i, \phi) \\ &= \sum_{i=1}^n l_i. \end{aligned}$$

Pour obtenir les équations de vraisemblance, nous calculons

$$\frac{\partial l_i}{\partial \beta_r} = \frac{\partial l_i}{\partial \theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_r}.$$

En utilisant (14.1), (14.4), (14.5) et (14.6), on obtient

$$\frac{\partial l_i}{\partial \beta_r} = \frac{(y_i - \mu_i)x_{ir}}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i}. \quad (14.7)$$

Pour maximiser la log-vraisemblance, on annule le système des p équations de vraisemblance égal à

$$\begin{cases} \sum_{i=1}^n \frac{(y_i - \mu_i)x_{i1}}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \\ \vdots \\ \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ip}}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \end{cases}$$

Ces équations n'étant en général pas des fonctions linéaires de β , il est nécessaire d'utiliser des méthodes itératives afin d'estimer $\hat{\beta}$.

Déterminons maintenant les termes de la matrice d'information de Fisher :

$$\begin{aligned} \mathbb{E}\left(\frac{\partial^2 l_i}{\partial \beta_r \partial \beta_s}\right) &= -\mathbb{E}\left[\left(\frac{\partial l_i}{\partial \beta_r}\right)\left(\frac{\partial l_i}{\partial \beta_s}\right)\right] \\ &= -\mathbb{E}\left[\frac{(y_i - \mu_i)x_{ir}}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{(y_i - \mu_i)x_{is}}{\text{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i}\right] \\ &= \frac{x_{ir}x_{is}}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2. \end{aligned} \quad (14.8)$$

En généralisant ce résultat à l'échantillon, on obtient :

$$\begin{aligned} a_{rs} &= -\mathbb{E}\left(\frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s}\right) \\ &= \sum_{i=1}^n \frac{x_{ir}x_{is}}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2. \end{aligned} \quad (14.9)$$

La matrice d'information de Fisher est donc de la forme

$$A = X'WX, \quad (14.10)$$

où W est une matrice diagonale d'éléments

$$w_i = \frac{1}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2. \quad (14.11)$$

14.1.2 Algorithmes

14.1.2.1 Algorithme de Newton-Raphson

Cet algorithme est basé sur le développement de Taylor, au second ordre et par rapport à β , du gradient de la log-vraisemblance. Soit $\beta^{(a)}$ la a^e approximation de $\hat{\beta}$ et considérons le développement de Taylor

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \beta} \Big|_{\beta^{(a)}} \\ &\approx \frac{\partial L}{\partial \beta} \Big|_{\beta^{(a)}} + \frac{\partial^2 L}{\partial \beta \partial \beta'} \Big|_{\beta^{(a)}} (\hat{\beta} - \beta^{(a)}). \end{aligned}$$

alors

$$\begin{aligned}\hat{\beta} - \beta^{(a)} &\approx \left[\left(-\frac{\partial^2 L}{\partial \beta \partial \beta'} \right)^{-1} \frac{\partial L}{\partial \beta} \right]_{\beta^{(a)}} \\ &= \delta^{(a)} \quad (\text{notation}).\end{aligned}$$

On peut ainsi construire une nouvelle valeur estimée

$$\beta^{(a+1)} = \beta^{(a)} + \delta^{(a)}.$$

$\delta^{(a)}$ peut constituer un critère d'arrêt en stoppant l'algorithme quand $\delta^{(a)}$ est suffisamment petit. Si l'on note $u^{(a)}$ le vecteur gradient et $H^{(a)}$ la matrice Hessienne calculés à la a^e itérations, on obtient la relation

$$\beta^{(a+1)} = \beta^{(a)} - (H^{(a)})^{-1} u^{(a)}.$$

14.1.2.2 Relation entre la méthode du scoring de Fisher et la méthode des moindres carrés pondérés itératifs (IRLS)

Dans l'algorithme de Fisher, la matrice Hessienne $H^{(a)}$ est remplacée par moins la matrice d'information de Fisher $A^{(a)}$:

$$\beta^{(a+1)} = \beta^{(a)} + (A^{(a)})^{-1} u^{(a)}.$$

En multipliant les deux termes de l'équation par $A^{(a)}$, on obtient

$$A^{(a)} \beta^{(a+1)} = A^{(a)} \beta^{(a)} + u^{(a)}. \quad (14.12)$$

En utilisant (14.9), la partie de droite de (14.12) devient

$$\sum_{i=1}^p \left[\sum_{i=1}^n \frac{x_{ir} x_{is}}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^{(a)^2} \right] \beta_s^{(a)} + \sum_{i=1}^n \frac{(y_i - \mu_i^{(a)}) x_{ir}}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^{(a)}, r = 1, \dots, p.$$

ce qui peut s'exprimer sous la forme

$$A^{(a)} \beta^{(a)} + u^{(a)} = X^t W^{(a)} z^{(a)},$$

où $W^{(a)}$ est W en (14.10) évalué en $\beta^{(a)}$, et $z^{(a)}$ est constitué des éléments

$$\begin{aligned}z_i^{(a)} &= \sum_{i=1}^p x_{ij} \beta_j^{(a)} + (y_i - \mu_i^{(a)}) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{(a)} \\ &= \eta_i^{(a)} + (y_i - \mu_i^{(a)}) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{(a)}, i = 1, \dots, n.\end{aligned} \quad (14.13)$$

En utilisant (14.10) pour $A^{(a)}$, (14.12) peut s'exprimer par

$$(X^t W^{(a)} X) \beta^{(a+1)} = X^t W^{(a)} z^{(a)}.$$

Ce sont les équations normales de la **méthode des moindres carrés pondérés** pour résoudre un modèle linéaire ayant comme variable dépendante $z^{(a)}$, comme variable indépendante la matrice X , et une matrice des poids $W^{(a)}$. La solution des équations est

$$\beta^{(a+1)} = (X^t W^{(a)} X)^{-1} X^t W^{(a)} z^{(a)}. \quad (14.14)$$

z est une forme généralisée de la fonction de lien $g(\mu)$ évaluée en y :

$$\begin{aligned} g(y_i) &\approx g(\mu_i) + (y_i - \mu_i)g'(\mu_i) \\ &= \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) \\ &= z_i . \end{aligned}$$

Ainsi, à chaque itération on calcule $z^{(a)}$ et $W^{(a)}$ pour obtenir une nouvelle estimation $\beta^{(a+1)}$ de β . Cette estimation permet de calculer un nouveau prédicteur linéaire $\eta^{(a+1)}$, et donc une nouvelle variable dépendante ajustée, ainsi que de nouveaux poids. D'où le nom de *méthode des moindres carrés pondérés itératifs*.

Asymptotiquement, l'inverse de la matrice d'information de Fisher constitue une estimation de la matrice de variance-covariance de $\hat{\beta}$, et par suite

$$\text{Cov}(\hat{\beta}) = (X^t \hat{W} X)^{-1} .$$

14.1.3 Simplification lors de l'utilisation d'un lien canonique

$$\theta_i = g(\mu_i) = \sum_{j=1}^n \beta_j x_{ij} .$$

En utilisant le fait que $\eta = X\beta$, on obtient

$$\begin{aligned} \frac{\partial \mu_i}{\partial \eta_i} &= \frac{\partial \mu_i}{\partial \theta_i} \\ &= \frac{\partial b'(\theta_i)}{\partial \theta_i} \\ &= b''(\theta_i) \\ &= \frac{\text{Var}(y_i)}{a(\phi)} . \end{aligned}$$

(14.7) devient

$$\frac{\partial l_i}{\partial \beta_j} = \frac{(y_i - \mu_i)x_{ij}}{a(\phi)} .$$

De plus, la matrice hessienne H est égale à moins la matrice d'information de Fisher : en effet, en utilisant (14.8),

$$\begin{aligned} \frac{\partial^2 l_i}{\partial \beta_r \partial \beta_s} &= \frac{(y_i - \mu_i)x_{ir}}{a(\phi)} \frac{(y_i - \mu_i)x_{is}}{a(\phi)} \\ &= \frac{\text{Var}(y_i)x_{ir}x_{is}}{a(\phi)^2} \end{aligned}$$

et

$$\begin{aligned} -\mathbb{E}\left(\frac{\partial^2 l_i}{\partial \beta_r \partial \beta_s}\right) &= \frac{x_{ir}x_{is}}{\text{Var}(y_i)} \left(\frac{\text{Var}(y_i)}{a(\phi)}\right)^2 \\ &= \frac{\text{Var}(y_i)x_{ir}x_{is}}{a(\phi)^2} . \end{aligned}$$

Ceci implique que les algorithmes de Newton-Raphson et du scoring de Fisher sont identiques.

14.1.4 Ajustement

- La log-vraisemblance vaut $l(\mu, y) = \log f(y, \theta)$; le critère d'ajustement est la **déviante pondérée**

$$D^*(y, \mu) = -2[l(\mu, y) - l(y, y)] ,$$

qui suit un χ^2 ;

- la **statistique de Pearson** :

$$X^2 = \frac{\sum (y - \hat{\mu})}{V(\hat{\mu})} ,$$

qui est la mesure d'ajustement de Pearson.

14.1.5 Étude des résidus

Deux types de résidus sont particulièrement utilisés :

- le **résidu de Pearson**, défini par

$$r_{Pi} = \frac{y - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} ;$$

la somme des carrés des résidus de Pearson est égale au χ^2 d'ajustement de Pearson ;

- le **résidu de la déviance**, défini par

$$r_{Di} = \text{signe}(y_i - \hat{\mu}_i) \sqrt{d_i} .$$

14.2 Données binaires

On note

$$\mathbb{P}(Y_i = 1) = \pi_i .$$

L'objectif est de rechercher la relation entre la probabilité de réponse $\pi = \pi(x)$, et les covariables $x = (x_1, \dots, x_p)$. On suppose que cette dépendance de π vis-à-vis des x_i est contenue dans la combinaison linéaire

$$\eta = g(\pi) = \sum_{j=1}^p x_j \beta_j .$$

À moins que des restrictions ne soient faites sur β , on a $-\infty < \eta < +\infty$. Aussi, étant donné que π est une probabilité, il faut une transformation $g(\pi)$ qui transforme l'intervalle $[0,1]$ en $]-\infty, +\infty[$. Trois fonctions sont usuellement employées :

- la **fonction logistique** :

$$g(\pi) = \log \left(\frac{\pi}{1 - \pi} \right) ;$$

- la **fonction probit** ou fonction inverse normale :

$$g(\pi) = \Phi^{-1}(\pi) ;$$

— la fonction log-log complémentaire :

$$g(\pi) = \log(-\log(1 - \pi)) .$$

Nota — Dans la cas du modèle logistique, on a bien que

$$\pi = \frac{\exp(\sum_j x_j \beta_j)}{1 + \exp(\sum_j x_j \beta_j)} .$$

La log-vraisemblance vaut

$$l(\pi; y) = \sum_{i=1}^n \left[y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + m_i \log(1 - \pi_i) \right] ,$$

où m_i est le nombre d'individus dans le groupe i .

$$\begin{aligned} \frac{\partial l}{\partial \beta_r} &= \sum_{i=1}^n \frac{y_i - m_i \pi_i}{\pi_i(1 - \pi_i)} \frac{d\pi_i}{d\eta_i} x_{ir} \\ \frac{\partial l}{\partial \beta} &= X^t(Y - \mu) . \end{aligned}$$

L'information de Fisher pour β vaut

$$\begin{aligned} -\mathbb{E}\left(\frac{\partial^2 l}{\partial \beta_r \partial \beta_s}\right) &= \sum_i \frac{m_i}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \beta_r} \frac{\partial \pi_i}{\partial \beta_s} \\ &= \sum_{i=1}^n \frac{m_i}{\pi_i(1 - \pi_i)} \left(\frac{\partial \pi_i}{\partial \eta_i}\right)^2 x_{ir} x_{is} \\ &= \{X^t W X\}_{rs} , \end{aligned}$$

en utilisant la forme matricielle vue en (14.10), avec

$$\begin{aligned} w_i &= \frac{m_i}{\pi_i(1 - \pi_i)} \left(\frac{\partial \pi_i}{\partial \eta_i}\right)^2 \\ &= m_i \pi_i(1 - \pi_i) . \end{aligned}$$

car

$$\begin{aligned} \frac{\partial \pi_i}{\partial \eta_i} x_{ir} &= \frac{\partial \pi_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r} \\ &= \frac{\partial \pi_i}{\partial \beta_r} \\ &= x_{ir} \frac{\exp(\sum_{j=1}^p x_{ij} \beta_j)}{[1 + \exp(\sum_{j=1}^p x_{ij} \beta_j)]^2} \\ &= x_{ir} (\pi_i(1 - \pi_i)) . \end{aligned}$$

14.2.1 Méthode itérative de Newton-Raphson

On se donne $\hat{\beta}_0$, et l'on calcule $\hat{\pi}_0$ et $\hat{\eta}_0$. On calcule alors, à partir de ces variables,

$$z_i = \hat{\eta}_i + \frac{y_i - m_i \hat{\pi}_i}{m_i} \frac{d\eta_i}{d\pi_i}.$$

Les estimateurs du maximum de vraisemblance vérifient

$$X^t W X \hat{\beta} = X^t W Z,$$

que l'on peut résoudre par itérations en utilisant la méthode standard des moindres carrés. On obtient

$$\hat{\beta}_1 = (X^t W X)^{-1} X^t W Z.$$

Propriétés 14.1 —

$$\begin{aligned} \mathbb{E}(\hat{\beta} - \beta) &\longrightarrow 0 && (n \rightarrow \infty), \\ \text{Cov}(\hat{\beta}) &\longrightarrow (X^t W X)^{-1} && (n \rightarrow \infty). \end{aligned}$$

La fonction de déviance vaut

$$\begin{aligned} D(y, \hat{\pi}) &= 2[l(\tilde{\pi}; y) - l(\hat{\pi}; y)] \\ &= 2 \sum_i \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i) \log\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right) \right\}. \end{aligned}$$

14.2.2 Méthode du scoring de Fisher

On constitue la variable dépendante ajustée (14.13)

$$\begin{aligned} z_i^{(a)} &= \eta_i^{(a)} + (y_i - \mu_i^{(a)}) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{(a)} \\ &= \eta_i^{(a)} + \frac{y_i - m_i \pi_i^{(a)}}{m_i \pi_i^{(a)} (1 - \pi_i^{(a)})}. \end{aligned}$$

Le système peut alors être résolu en utilisant (14.14).

14.3 Modèle linéaire généralisé à effets mixtes

14.3.1 Définition

Un GLM à effets mixtes peut se définir à partir d'un GLM de la façon suivante. Supposons que l'on ait K observations (y_1, \dots, y_K) de Y , telles que

$$Y = \mu + e,$$

où e est un vecteur de termes d'erreur de moyenne nulle et de matrice de variance-covariance V . Considérons la part systématique $\eta = g(\mu)$ du GLM, et définissons-la comme étant égale à

$$\eta = X\beta + B_1b_1 + \cdots + B_nb_n, \quad (14.15)$$

où :

- η est un vecteur de dimension $K \times 1$;
- X est la matrice de dimension $K \times p$ des covariables dont les valeurs sont connues ;
- β est un vecteur inconnu d'effets fixes, de dimension $p \times 1$;
- $B_i, i=1, \dots, n$, est une matrice connue de dimension $K \times q_i$;
- $b_i, i=1, \dots, n$, est un vecteur inconnu d'effets aléatoires et de dimension $q_i \times 1$.

14.3.2 Estimation des paramètres

Contrairement aux GLM traditionnels ou aux modèles à effets mixtes linéaires, il n'existe pas de méthode « standard » dans ce contexte. Nous détaillerons l'approche de Anderson et Aitkin.

Considérons le modèle (14.15) où les B_i sont des vecteurs de dimension $K \times 1$ dont toutes les composantes sont nulles sauf de la $(\sum_{j=1}^{i-1} k_j + 1)^{\text{e}}$ à la $(\sum_{j=1}^i k_j)^{\text{e}}$ composante, et les b_i sont des scalaires. Il est clair que $\sum_{i=1}^n k_i = K$. Dans le contexte des mesures répétées, en utilisant les mêmes notations que dans la partie 3, le prédicteur η_{it} , c'est-à-dire du i^{e} sujet au temps t s'écrit

$$\eta_{it} = \left[\sum_{j=1}^p x_{itj} \beta_j \right] + b_i.$$

Il est clair que ce modèle comporte un seul effet aléatoire qui est constant pour un individu donné. La conséquence est que, conditionnellement à b_i , les observations y_{it} sont indépendantes. Ces modèles sont dénommés **modèles avec ordonnée à l'origine aléatoire**.

On peut remarquer que si b est distribué suivant une loi normale $\mathcal{N}(0, \sigma^2)$ où σ^2 représente la composante de variance associée à b , alors le **coefficient de corrélation intra-classe** ρ est

$$\rho = \frac{\sigma^2}{1 + \sigma^2}.$$

La matrice D est alors bloc-diagonale avec des sous-matrices D_i de type *exchangeable correlation*. Notons de plus que $b_i = \sigma a_i$ où a est distribué suivant une loi normale centrée réduite. Dans ce cas, le prédicteur du i^{e} sujet au temps t s'écrit

$$\eta_{it} = \left[\sum_{j=1}^p x_{itj} \beta_j \right] + \sigma a_i.$$

La log-vraisemblance du modèle s'exprime alors comme suit :

$$\tilde{l}(\beta, \sigma) = \sum_{i=1}^n \log \left\{ \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{k_i} f(y_{it}; \beta, \sigma) \right] v(a_i) da_i \right\}, \quad (14.16)$$

où $v(a_i)$ est la fonction de densité d'une loi normale centrée réduite.

Anderson et Aitkin montrent que les paramètres β et σ peuvent être estimés par l'*EM algorithm*. Pour utiliser cet algorithme, il est nécessaire de définir la log-vraisemblance complète, c'est-à-dire en supposant que a est connu. La log-vraisemblance complète est alors

$$l(\beta, \sigma) = \sum_{i=1}^n \sum_{t=1}^{k_i} \log [f(y_{it}; \beta, \sigma) v(a_i)] . \quad (14.17)$$

L'*EM algorithm* est un algorithme itératif constitué de deux phases exécutées alternativement :

- la première phase est la phase d'estimation de l'algorithme (*E-step*) où est estimée non pas la vraisemblance complète, mais l'espérance de celle-ci, et conditionnellement aux données observées et aux estimations courantes des paramètres du modèle ;
- la seconde est la phase de maximisation de l'algorithme (*M-step*) qui consiste à trouver les quantités $\hat{\beta}$ et $\hat{\sigma}$ qui maximisent l'espérance de $l(\beta, \sigma)$.

En pratique, l'algorithme nécessite de résoudre des intégrales du type $\int_{-\infty}^{+\infty} f(\cdot) v(a) da$. Anderson et Aitkin proposent d'utiliser une procédure d'intégration numérique par quadrature de Gauss. La procédure nécessite de se fixer un nombre q de points d'intégration. On peut alors obtenir à partir de tables ou de routines les coordonnées a_q et les pondérations A_q utilisées dans l'intégration numérique.

Sixième partie

**ÉQUATIONS D'ESTIMATION
GÉNÉRALISÉES**

Quasi-vraisemblance

15.1 Vraisemblance marginale

Il s'agit d'éliminer les paramètres de nuisance. Si θ est le paramètre d'intérêt et β celui de nuisance, on élimine β de la vraisemblance en travaillant avec l'ensemble de contrastes

$$\begin{aligned} R &= (I - P_X)Y \\ &= (I - X(X^tX)^{-1}X^t)Y, \end{aligned}$$

de moyenne nulle et dont la distribution ne dépend pas de β .

15.2 Vraisemblance conditionnelle

On utilise la densité conditionnelle de Y sachant le paramètre d'intérêt.

15.3 Quasi-vraisemblance

On suppose que les composantes du vecteur Y sont indépendantes, de moyenne μ et de matrice de covariance $\sigma^2V(\mu)$, où σ^2 est inconnu et $V(\mu)$ connue. Le paramètre d'intérêt β se rattache à la dépendance de μ vis-à-vis des covariables x . Peu importe la nature de cette relation : nous noterons simplement $\mu(\beta)$. σ^2 est supposé constant — *i.e.* ne dépendant pas de β . Puisque les composantes de Y sont supposées indépendantes, la matrice $V(\mu)$ doit être diagonale :

$$V(\mu) = \text{diag}\{V_1(\mu), \dots, V_n(\mu)\}.$$

On suppose de plus que $V_i(\mu)$ ne dépend que de la i^{e} composante de μ :

$$V(\mu) = \text{diag}\{V_1(\mu_1), \dots, V_n(\mu_n)\}.$$

On considère une unique composante de Y . D'après ce qui précède, la fonction

$$\begin{aligned} U &= u(\mu; Y) \\ &= \frac{Y - \mu}{\sigma^2 V(\mu)} \\ &= \sum_{i=1}^n \frac{y_i - \mu_i}{\sigma^2 V(\mu_i)}. \end{aligned}$$

a les propriétés de log-vraisemblance (14.3) et (14.4) :

$$\begin{aligned} \mathbb{E}(U) &= 0, \\ \text{Var}(U) &= \frac{1}{\sigma^2 V(\mu)}, \\ -\mathbb{E}\left(\frac{\partial U}{\partial \mu}\right) &= \frac{1}{\sigma^2 V(\mu)}. \end{aligned}$$

Définition 15.1 — *L'intégrale*

$$Q(\mu; y) = \int_y^\mu \frac{y - t}{\sigma^2 V(t)} dt,$$

si elle existe, est la **fonction de quasi-vraisemblance** de μ , basée sur la donnée y . C'est en réalité la fonction de log-quasi-vraisemblance.

Puisque les composantes de Y sont indépendantes, la quasi-vraisemblance complète vaut

$$Q(\mu; y) = \sum_{i=1}^n Q_i(\mu_i; y_i).$$

Définition 15.2 — *La fonction de quasi-déviance est*

$$\begin{aligned} D(y; \mu) &= -2\sigma^2 [Q(\mu; y) - Q(y; y)] \\ &= -2 \int_y^\mu \frac{y - t}{V(t)} dt, \end{aligned}$$

qui est indépendante de σ^2 .

L'objectif est de maximiser Q , ou encore d'annuler les dérivées premières $U(\beta)$ de Q par rapport à β . Il est nécessaire de calculer

$$\begin{aligned} \frac{\partial Q_i}{\partial \beta_r} &= \frac{\partial Q_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_r} \\ &= U_i D_{ir}, \end{aligned}$$

avec D matrice $n \times p$, d'éléments $D_{ir} = \frac{\partial \mu_i}{\partial \beta_r}$.

Exprimé sous forme matricielle, le système à résoudre est de la forme

$$U(\beta) = \frac{1}{\sigma^2} D^t V(\mu)^{-1} (Y - \mu), \quad (15.1)$$

qui est appelée **fonction de quasi-score**.

La matrice de covariance de $U(\beta)$, qui est aussi $\mathbb{E}\left(\frac{\partial U(\beta)}{\partial \beta}\right)$, est

$$i_\beta = \frac{1}{\sigma^2} D^t V^{-1} D. \quad (15.2)$$

Pour les fonctions de quasi-vraisemblance, cette matrice joue le même rôle que l'information de Fisher pour les fonctions de vraisemblance ordinaire.

Théorème 15.1 — *On suppose que :*

- (i) *la dérivée troisième de $\mu(\beta)$ existe;*
- (ii) *les 3 premiers moments de la distribution de Y existent;*
- (iii) *i_β/n converge vers une matrice définie positive quand n tend vers l'infini.*

Alors

$$\hat{\beta} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\beta, \frac{1}{i_\beta}\right).$$

Théorème 15.2 — *Soient deux hypothèses H_A et H_B emboîtées, avec $\dim A < \dim B$. Alors, sous H_A , la différence en déviance*

$$D(\hat{\mu}_B, \hat{\mu}_A) = D(y, \hat{\mu}_A) - D(y, \hat{\mu}_B)$$

suit asymptotiquement une loi du χ^2 à $B - A$ degrés de liberté.

15.4 Méthode de Newton-Raphson

Commençant avec une valeur arbitraire $\hat{\beta}_0$ suffisamment proche de $\hat{\beta}$, la méthode de Newton-Raphson conduit à

$$\hat{\beta}_1 = \hat{\beta}_0 + (\hat{D}_0^t \hat{V}_0^{-1} \hat{D}_0)^{-1} \hat{D}_0^t \hat{V}_0^{-1} (y - \hat{\mu}_0).$$

15.5 Méthode de Fisher

Pour estimer les $\hat{\beta}$, on utilise l'algorithme de Fisher, ce qui donne l'expression

$$\begin{aligned} \beta^{(a+1)} &= \beta^{(a)} + \frac{1}{i_\beta^{(a)} u^{(a)}} \\ &= \beta^{(a)} + (D^{(a)t} V^{(a)-1} D^{(a)})^{-1} D^{(a)t} V^{(a)-1} (y - \mu^{(a)}), \end{aligned} \quad (15.3)$$

laquelle peut s'exprimer sous la forme

$$(D^{(a)t} W^{(a)} D^{(a)}) \beta^{(a+1)} = (D^{(a)t} W^{(a)} Z^{(a)}), \quad (15.4)$$

où $W^{(a)} = V^{(a)-1}$ et $Z^{(a)}$ est une variable dépendante ajustée égale à

$$D^{(a)}\beta^{(a)} + (y - \mu^{(a)}) .$$

On peut remarquer que σ^2 n'intervient pas dans l'estimation des $\hat{\beta}$. L'estimation de σ^2 ne peut se faire par un calcul de vraisemblance ; il est généralement estimé directement sur l'échantillon à partir de la statistique de Pearson généralisée

$$\tilde{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \frac{y_i - \hat{\mu}_i}{V_i(\hat{\mu}_i)} . \quad (15.5)$$

15.6 Conditions d'application

Le concept de quasi-vraisemblance est utilisé dans deux types de situation :

- l'étude de modèles pour lesquels la connaissance de la distribution de Y se limite aux deux premiers moments (en particulier lorsque $V(\mu) = 1$ – variance constante – ou lorsque $V(\mu) = \mu^2$ – coefficient de variation constant) ;
- l'extension de la famille exponentielle naturelle par l'introduction d'un **paramètre de supra-dispersion**.

Mais le concept de quasi-vraisemblance exclue la prise en compte de corrélation entre les observations. D'où l'introduction, par Liang et Zeger (1986), du concept d'**équations d'estimation généralisées (GEE)**, qui sont la généralisation de la notion de quasi-vraisemblance à des observations dépendantes.

Une autre limitation de la méthode de quasi-vraisemblance est que la forme de la fonction de variance est supposée connue. Une extension, appelée **quasi-vraisemblance étendue (extended quasi-likelihood)**, a été proposée par Nelder (1987). Dans ce modèle, la fonction de variance est paramétrée. Les propriétés de cette méthode ont été récemment étudiées par simulation (Nelder, 1992).

Équations d'estimation généralisées

16.1 Modèle

On note y_{it} la réponse observée chez le i^e sujet au temps t , et x_{itj} la valeur de la j^e covariable mesurée chez le i^e sujet au temps t . Nous supposons que l'échantillon est constitué de n sujets et que l'on observe p covariables aux temps k_i . La dimension de Y est donc égale à $K = \sum_i k_i$, et celle de X est égale à $K \times p$.

Les **équations d'estimation généralisées (GEE)** permettent de modéliser l'espérance marginale de y_{it} , soit $\mathbb{E}(y_{it}) = \mu_{it}$. C'est une méthode qui fournit des estimations « moyennées » sur la population. En utilisant le même raisonnement que pour la quasi-vraisemblance, définissons la variance de y_{it} et la fonction de lien reliant μ_{it} aux covariables :

$$\begin{aligned}\text{Var}(y_{it}) &= \phi \nu^*(\mu_{it}) \\ g^*(\mu_{it}) &= X^t \beta^* .\end{aligned}$$

β^* mesure l'effet d'une covariable sur la réponse moyenne au niveau de la population, et non un effet individuel.

Notons

$$\mu_i = \{g^{*-1}(x'_{i1}\beta^*), \dots, g^{*-1}(x'_{ik_i}\beta^*)\}^t$$

et notons A_i une matrice diagonale de dimension $k_i \times k_i$ dont les éléments diagonaux sont constitués par les $\nu^*(\mu_i)$. Sous l'hypothèse d'indépendance des observations chez le même sujet,

$$\text{Cov}(y_i) = \phi A_i .$$

Le plus souvent, cette hypothèse n'est guère soutenable ; on définit alors une **matrice de corrélation, dite « de travail »** $R_i \alpha$ dépendant d'un vecteur α de paramètres inconnus. Pour estimer β^* , Liang et Zeger proposent de résoudre un système d'équations analogues aux équations de quasi-vraisemblance (15.1) :

$$U(\beta^*) = \sum_{i=1}^n D_i V_i^{-1}(\alpha)(y_i - \mu_i) = 0 \quad (16.1)$$

où

$$D_i = \frac{\partial \mu_i'}{\partial \beta^*}$$

et

$$V_i(\alpha) = \phi \sqrt{A_i} R_i(\alpha) \sqrt{A_i}.$$

Liang et Zeger montrent que, sous les conditions d'une spécification correcte de μ_i et les conditions usuelles de régularité, $\hat{\beta}^*$ est un **estimateur consistant et asymptotiquement gaussien** de β^* ($n \rightarrow \infty$). En particulier, ces propriétés sont respectées même en cas de mauvaise spécification de V_i .

16.2 Estimation des paramètres

Pour estimer les paramètres, on alterne une phase d'estimation de β^* fondé sur l'algorithme de Fischer, et une phase d'estimation de α et ϕ par la méthode des moments.

16.2.1 Estimation de β^*

En utilisant les valeurs courantes des estimations $\alpha^{(a)}$ et $\phi^{(a)}$, on en déduit en utilisant une démarche analogue à (15.2) que

$$\beta^{*(a+1)} = \beta^{*(a)} + \left\{ \sum_{i=1}^n D_i^{(a)t} [V_i^{(a)}(\alpha)]^{-1} D_i^{(a)} \right\} \left\{ \sum_{i=1}^n D_i^{(a)t} [V_i^{(a)}(\alpha)]^{-1} (y_i - \mu_i^{(a)}) \right\}. \quad (16.2)$$

Dans le cas où $k_1 = \dots = k_n = k$, en s'inspirant de (15.3), l'équation ci-dessus peut s'exprimer sous forme matricielle

$$(\Delta^{(a)t} W^{(a)} \Delta^{(a)}) \beta^{(a+1)} = \Delta^{(a)t} W^{(a)} Z^{(a)}. \quad (16.3)$$

16.2.2 Estimations de α et ϕ

Pour estimer α et ϕ , on utilise, comme dans le cas de la quasi-vraisemblance, les résidus de Pearson définis pour le i^e individu par

$$\hat{r}_i = \frac{y_i - \hat{\mu}_i}{\sqrt{A_i}}. \quad (16.4)$$

ϕ se définit de façon analogue à (15.4) par

$$\hat{\phi} = \frac{\sum_{i=1}^n \hat{r}_i' \hat{r}_i}{\sum_{i=1}^n k_i}. \quad (16.5)$$

Pour estimer α , l'approche générale consiste à utiliser des fonctions simples concernant les termes de covariance des résidus, de la forme

$$\hat{R}_{uv} = \sum_{i=1}^n \frac{\hat{r}_{iu} \hat{r}_{iv}}{n-p}.$$

Plusieurs formes de matrices de variance-covariance peuvent être spécifiées, la plus simple étant la matrice identité (dans ce dernier cas, l'estimation de β est identique à celle de l'estimation sous hypothèse d'indépendance des observations, à l'exception toutefois de la variance. . . ceci sera vu un peu plus loin). Une autre possibilité est de supposer la matrice de corrélation connue et d'en spécifier les coefficients. À l'opposé, on peut considérer la matrice de corrélation comme inconnue et estimer ses composantes qui sont de la forme

$$\hat{R}_{uv}(\alpha) = \frac{\hat{R}_{uv}}{\hat{\phi}} .$$

Nous verrons plus loin quelques formes courantes de matrices de corrélation.

16.2.3 Estimation de la variance de $\hat{\beta}^*$

En utilisant (15.2), une estimation naïve de la variance de β^* est fournie par

$$\text{Cov}(\hat{\beta}^*) = \frac{\hat{\phi}}{\sum_{i=1}^n \Delta_i^{(a)t} [V_i^{(a)}(\alpha)]^{-1} \Delta_i^{(a)}} .$$

Liang et Zeger montrent que

$$V_{\hat{\beta}^*} = M_0^{-1} M_1 M_0^{-1} , \quad (16.6)$$

où

$$M_0 = \sum_{i=1}^n \hat{\Delta}_i^t \hat{V}_i^{-1} \hat{\Delta}_i ,$$

$$M_1 = \sum_{i=1}^n \hat{\Delta}_i^t \hat{V}_i^{-1} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)^t \hat{V}_i^{-1} \hat{\Delta}_i .$$

$V_{\hat{\beta}^*}$ est consistante même lorsque $\text{Cov}(y_i) \neq V_i$.

16.3 Différentes matrices de travail $R(\alpha)$

Quelques matrices de corrélation sont présentées en fin de chapitre.

► La première famille est la famille des matrices de corrélation non stationnaire d'ordre m . Elles s'écrivent

$$R(\alpha)_{uv} = \begin{cases} 1 & \text{si } u = v \\ \alpha_{uv} & \text{si } |u - v| \leq m \\ 0 & \text{si } |u - v| > m \end{cases}$$

Chaque $\hat{\alpha}_{ij}$ peut s'exprimer par

$$\hat{\alpha}_{ij} = \sum_{i=1}^n \frac{\hat{r}_{iu} \hat{r}_{iv}}{\hat{\phi}(n-p)} .$$

► La deuxième famille est celle des matrices de corrélation stationnaires d'ordre m . Notons $t = |u - v|$. La matrice de corrélation est ici égale à

$$R(\alpha)_{uv} = \begin{cases} 1 & \text{si } t = 0 \\ \alpha_t & \text{si } t \leq m \\ 0 & \text{si } t > m \end{cases}$$

et une estimation de $\hat{\alpha}_{ij}$ peut s'exprimer par

$$\hat{\alpha}_t = \sum_{u=1}^{k-t} \frac{\hat{\alpha}_{u,u+t}}{k-t}.$$

► Une autre possibilité est de considérer α comme étant le même pour tout couple (u,v) , $u \neq v$ (cas d'une *exchangeable correlation*). Liang et Zeger proposent d'estimer α par

$$\hat{\alpha} = \frac{\sum_{i=1}^n \sum_{u>v} \hat{r}_{iu} \hat{r}_{iv}}{\hat{\phi} \sum_{i=1}^n \frac{1}{2} k_i (k_i - 1) - p}.$$

Toutes ces matrices peuvent s'exprimer sous la forme

$$R(\hat{\alpha}) = \frac{T(\hat{\alpha})}{\hat{\phi}}, \quad (16.7)$$

où $T(\alpha)$ est une matrice qui ne dépend pas de ϕ . La conséquence en est alors que le terme ϕ disparaît dans l'expression de V_i , ce qui entraîne que les estimations de $\hat{\beta}^*$ et $\text{Var}(\hat{\beta}^*)$ ne dépendent plus de ϕ .

► La dernière famille de matrices de corrélation est constituée par les matrices traduisant une corrélation autorégressive d'ordre 1. La corrélation entre deux mesures est alors de la forme

$$\alpha^{|u-v|}.$$

On peut estimer α par la moyenne des coefficients de corrélation calculés sur chaque série. Il faut noter que dans cette situation, $R(\hat{\alpha})$ n'est pas décomposable suivant (16.7), car il fait intervenir $\hat{\phi}$ à la puissance $-|u-v|$. En revanche, cette situation s'accommode bien d'un nombre variable de mesures, ainsi que d'intervalles non constants entre les mesures.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} & \alpha_{25} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} & \alpha_{35} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} & 1 & \alpha_{45} \\ \alpha_{15} & \alpha_{25} & \alpha_{35} & \alpha_{45} & 1 \end{pmatrix}$$

independence

unstructured

$$\begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & 0 & 0 \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} & 0 \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} & \alpha_{35} \\ 0 & \alpha_{24} & \alpha_{34} & 1 & \alpha_{45} \\ 0 & 0 & \alpha_{35} & \alpha_{45} & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & \alpha_1 & \alpha_2 & 0 & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 & 0 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 & \alpha_2 \\ 0 & \alpha_2 & \alpha_1 & 1 & \alpha_1 \\ 0 & 0 & \alpha_2 & \alpha_1 & 1 \end{pmatrix}$$

not stationnary (order 2)

stationnary (order 2)

$$\begin{pmatrix} 1 & \alpha & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & \alpha & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 \\ \alpha & 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha^2 & \alpha & 1 & \alpha & \alpha^2 \\ \alpha^3 & \alpha^2 & \alpha & 1 & \alpha \\ \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}$$

exchangeable correlation *autoregressive (order 1)*

FIGURE 16.1 — Quelques matrices de travail.

16.4 Extensions des GEE

Plusieurs développements autour de la méthodologie GEE ont été proposés. Thall et Vail (1990) et Paik (1992) ont développé des modèles où la matrice de variance-covariance peut être paramétrée par des covariables. Paik montre que l'ignorance d'une hétérogénéité de la variance (par exemple un phénomène d'hétéroscédasticité en fonction du temps) se traduit par une perte d'efficacité pour l'estimation des β .

Rotnitzky et Jewell (1990) construisent des tests de signification de type *test du score* ou *test de Wald* dans le contexte des GEE. Ils proposent également un test ajusté basé sur la déviance calculée sous l'hypothèse d'indépendance des observations.

Citons enfin l'article de Zeger, Liang et Albert (1988) qui expose un modèle de type *subject-specific* à partir de la méthodologie des GEE.